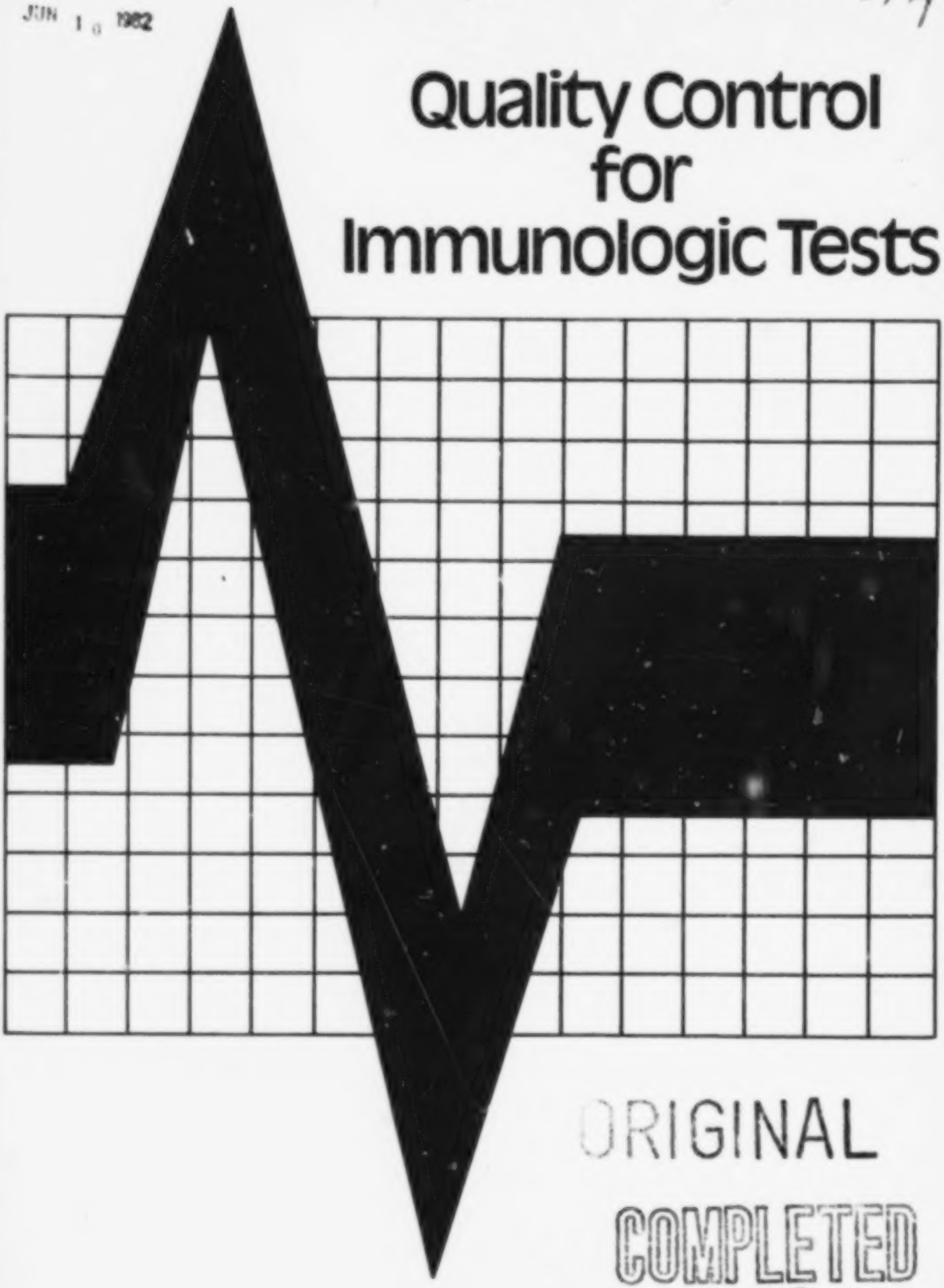
JUN 18 1982

HE [illegible]

Quality Control for Immunologic Tests



ORIGINAL

COMPLETED

128

U.S. Department of Health, Education, and Welfare / Public Health Service / Center for Disease Control

QUALITY CONTROL FOR IMMUNOLOGIC TESTS

AUTHORS: Roger N. Taylor, Ph.D.
Alan Y. Huong, Ph.D.
Karen M. Fulford, M.S.
Vincent A. Przybyszewski, M.S.
Thomas L. Hearn, M.S.

JUNE 1979

U.S. DEPARTMENT OF HEALTH, EDUCATION, AND WELFARE
PUBLIC HEALTH SERVICE
CENTER FOR DISEASE CONTROL
BUREAU OF LABORATORIES
ATLANTA, GEORGIA 30333

HEW Publication No. (CDC) 79-8376

I

BLANK PAGE

PREFACE

This monograph is intended to be a general guideline to bench level technologists in establishing and maintaining an effective and efficient quality control program. It is intended as a supplement to the training manuals which indicate specific quality control procedures for specific tests. The purpose is to stimulate quality control consciousness and thinking and to provide a source of references for additional literature, methods, and standards.

The section dealing with quality control calculations has been developed to a greater extent than other sections because such information is not readily available. Even though many may not be familiar with these tests and calculations, we hope that they will learn to do them and use them in their quality control systems.

The tests in this publication include a wide range of commonly used immunologic tests. Some of the more specialized tests, especially the more recently developed cellular immunity tests, have not been included.

We wish to acknowledge the contributions of the following people in preparing and reviewing the manuscript:

Cynthia E. Bramblett
Helen L. Casey, Ph.D.
Samuel P. Caudill, Ph.D.
William B. Cherry, Ph.D.
Elliott Churchill, M.A.
Glenda S. Cowart, M.S.
Nancy J. Evans, M.T. (ASCP)
David S. Gordon, M.D.
William K. Harrell, Ph.D.
Wallis L. Jones, Ph.D.
Gayle D. McClary, M.T. (ASCP)
R. Michael Parvin, M.T. (ASCP)
Alwilda L. Wallace, M.S.
Rhonda S. Whalen, M.T.(ASCP)

BLANK PAGE

CONTENTS

Page

I. **Introduction** 1

II. **General Quality Control Functions** 4
  A. Prodedure Manual 4
  B. Selection of Test or Procedure 4
  C. Collection of Specimens 5
  D. Control and Reference Sera 6
    1. Source 6
    2. Preparation 6
    3. Storage 7
    4. Reference Materials 7
  E. Reagents 7
  F. Equipment and Instruments 8
  G. Performance of Tests 12
  H. Reporting and Record Keeping 12
  I. Implementation of Quality Control 13

III. **Quality Control of Specific Procedures** 15
  A. Precipitin Tests 15
    1. Principles Involved 15
    2. Problems Frequently Encountered 16
    3. Control Recommendations 16
    4. Standardization 17
  B. Agglutination Tests 17
    1. Principles Involved 17
    2. Problems Frequently Encountered 18
    3. Control Recommendations 19
    4. Standardization 20
  C. Complement Fixation Tests 20
    1. Principles Involved 20
    2. Problems Frequently Encountered 21
    3. Control Recommendations 22
    4. Standardization 22

Contents

Page

D. Inhibition Tests. 23
    1. Principles Involved. 23
    2. Problems Frequently Encountered 23
    3. Control Recommendations 25
    4. Standardization. 25
E. Immunofluorescence Microscopy 26
    1. Principles Involved. 26
    2. Problems Frequently Encountered 27
    3. Control Recommendations 28
    4. Standardization. 29
F. Radioimmunoassay 29
    1. Principles Involved. 29
    2. Problems Frequently Encountered 30
    3. Control Recommendations 31
G. Enzyme-Immunoassay 34
    1. Principles Involved. 34
    2. Problems Frequently Encountered 35
    3. Control Recommendations 36
H. Electrophoresis 36
    1. Principles Involved. 36
    2. Problems Frequently Encountered 37
    3. Control Recommendations 38

I. Other Tests

    1. Ox Cell Hemolysin (OCH) Test for Infections Mononucleosis 38
        a. Principles Involved. 38
        b. Problems Frequently Encountered 39
        c. Control Recommendations 39
        d. Standardization. 39

IV. Standardization. 40

V. External Quality Control 51

VI. Quality Control Calculations 54
    A. Introduction to Statistical Calculations with Geometric Data. 55

Contents

Page

B. Tests of Goodness of Fit 63
  1. .Chi-Square Test 63
  2. Kolmogorov-Smirnov (K-S) Test 67
C. Common Statistical Tests 67
  1. Determining Confidence Intervals 67
  2. Determination of Sample Size 69
  3. Student's t Test 70
    a. Two Group t Test 72
    b. Paired t Test 73
  4. Analysis of Variance 74
  5. Components of Variance Analysis 76
  6. Nonparametric Tests 78
    a. Sign Test 79
    b. Fisher Exact Probability Test 80
    c. Chi-Square Test 82

Bibliography 87

Glossary of Terms 95

Appendices 101
  A. Cumulative Normal Frequency Distribution 103
  B. Cumulative Distribution of Chi-Square 104
  C. Critical Values in the K-S Test 105
  D. Distribution of t 107
  E. Distribution of F 108

Page

Tables

Table 1 Suggested Monitoring of the Temperature of Laboratory Equipment .9
Table 2 Suggested Monitoring of Laboratory Equipment 10
Table 3 List of Available Standards 41
Table 4 Products for which Performance Specifications Have Been Established 43
Table 5 Log Base Conversion Table 59
Table 6 CDC Proficiency Testing Data for IgG 66
Table 7 IgG Results Compared by Antibody Source 77
Table 8 Appropriate Tests and Calculations for Statistical Analyses 84

Figures

Figure 1 Immunology Quality Control Chart 60
Figure 2 Distribution of Test Results. 62

I. INTRODUCTION

According to Murphy's Law, if anything can go wrong it will. The purpose of quality control is to prevent as many errors as possible and to detect those that do occur. Total quality control means that every variable that could possibly affect the quality of the test results has been controlled. Obviously such a situation is ideal rather than reasonable. However, a practical quality control system should control most of the factors that are likely to affect the test results. Determining which controls are important components of a quality control system, how often to include the controls, how many controls to include, and how to use the control results requires experience and knowledge. This monograph provides information and recommendations to laboratorians so they can intelligently design quality control systems which are appropriate for their circumstances.

Numerous sophisticated quality control recommendations are available for other clinical laboratory disciplines, but those for immunologic tests have generally been neglected. Considerable effort has been expended in developing and publishing quality control manuals for chemistry, hematology, and other clinical laboratory specialties, but the efforts in immunology have been limited and fragmentary. The fact that the distribution of results from most immunologic tests is more nearly log-normal than normal is seldom mentioned, and the implications and consequences of this fact are almost always disregarded. The application of statistical methods to immunologic data is particularly limited. We hope that this monograph provides a starting point for the development of an increased quality control consciousness among immunologists and thus a catalyst for the development and implementation of better quality control methods and programs.

This manual was developed as the result of several stimuli. One was the general atmosphere of concern about the validity of the results produced in clinical laboratories. Another factor was our experience with the results of proficiency testing programs in clinical laboratories, which includes the observations that laboratories in proficiency testing programs frequently report that they do not use the recommended control procedures, that they deviate from standardized methods, that they fail to detect product and reagent deficiencies, and that they even report test results when they have also reported that the test system is "out of control" (*108,110,111*).

In the past few years, serologic and immunologic tests have been more widely used in the clinical laboratory. As new methods and techniques are developed and adapted, adequate quality control must be developed to assure the reliability and usefulness of clinical laboratory test results. No single quality control system is appropriate for every laboratory. Such variables as the volume of testing, the qualifications and interests of the personnel, the composition of the patient population, the type of laboratory, the purpose for which the results will be used, budgetary limitations, and external factors must all be considered. The implementation of an effective program requires knowledge and experience in order to achieve the proper balance between ideal quality control requirements and the limiting factors. Since variation in test results is a composite of the variation of each step of the total procedure, prime consideration should be given to those components which contribute most to the total variation. The components of variation which are most amenable to improvement should be most carefully examined.

An effective quality control program should cover all aspects of the laboratory. Of greatest importance are personnel with adequate training and experience, since much of the quality control is a function of professional judgment (*106*). Proper specimen collection and processing of results are as important as the analytical procedure itself. Reagents must be of good quality, technique must be adequate to maintain precision and accuracy, and there must be methods for detecting errors and taking corrective steps when analyses are out of control. Preventive maintenance, continuous training of employees, documentation, and coordination of activities are also part of a comprehensive quality control program (*6,92*).

Improved accuracy and precision, smoother laboratory operation, and better morale may follow the implementation of a quality control program which is easy to maintain and interpret. When possible, the program should monitor performance quantitatively rather than subjectively and should cost no more than absolutely necessary. The major purpose of the quality control program is to confirm that results are of high quality and that the physician can rely on them.

Some aspects of quality control concern prevention, whereas others concern verification of following proper procedures. Using properly evaluated and prepared reagents, maintaining and calibrating equipment, checking temperatures, and applying proper techniques are preventive measures. Control sera and reaction mixtures are used to verify technique and detect errors. When results indicate that a procedure is "out of control," the source of error must be determined and the problem corrected as efficiently as possible.

Whether it is better to have each technologist involved in quality control or to have a "quality control technologist" is a debatable question, but if a specific person is given the responsibility for establishing an effective quality control program, his or her rsponsibility should include those of collecting and summarizing data and the rapid and effective transfer of information to the laboratory director (*7,93,131*). Records of quality control that reflect proficiency testing results, routine control sample results, equipment maintenance, calibration, and reagent testing should be maintained.

Procedural manuals should be available at the bench for each procedure offered by the laboratory. Any changes in procedures should be dated and initialed in the manual by the laboratory director or his representative.

II. GENERAL QUALITY CONTROL FUNCTIONS

A. PROCEDURE MANUAL

An important element in maintaining day-to-day uniformity in laboratory results is an established procedures manual which details all phases of the laboratory's operation (including safety precautions) and is used by all laboratory personnel. It should include instructions for collecting, transporting, and storing specimens, for preparing and storing reagents, and for performing tests. In addition, the controls and calibrators to be used should be listed along with directions for their use, expected results, and instructions for corrective measures if the expected results are not obtained. Although attempting to develop new methods and trying new methods described in the literature are to be encouraged, procedures used to test patient specimens should be changed only with the approval of the director or designee after comparative studies in the laboratory have shown that the new ones are satisfactory (*17*).

B. SELECTION OF TEST OR PROCEDURE

As new tests and methods are developed for various analytes (antibodies or antigens), the most appropriate must be chosen for each laboratory's needs. A number of factors must be considered including bias, specificity, sensitivity, precision, cost, and ease of performance. Bias, specificity and sensitivity may be related. Frequently the more sensitive a test, the less specific it is (*106*). Bias may result from low specificity or sensitivity. The population to be tested may influence the decision of what test to use; e.g., screening essentially normal patients does not necessarily require the same methods appropriate for following patients with a confirmed disease.

To determine the presence of bias, the proposed method should be compared with other reliable methods, preferably with a standard method or clinical data. The same specimens should be run with both methods in the same laboratory and the results compared, although interlaboratory comparisons are also useful. If the results from the different methods do not agree one must determine the reason for the difference and then decide which result is more useful.

The specificity of a method is evaluated by testing negative samples and samples containing substances which might cause interference. Closely related or cross-reacting substances frequently found in clinical specimens should be included.

The sensitivity of the method being evaluated should be compared to that of other methods, but the purpose of the test must also be considered. In general, a definitive test need not be as sensitive as a screening test. The test should distinguish between normal and abnormal levels of analyte.

The precision of a quantitative or semiquantitative test must be evaluated in light of the precision required for the clinical application of the test results. Many factors affect precision, but one that is frequently overlooked in serologic tests is the size of the dilution increments. If all other variables are held constant, serologic tests tend to become less precise as the size of the dilution increment increases. For example, it should be expected that a test based on a 4-fold dilutions would be less precise than the same test with 2-fold dilutions.

The predictive value of a positive test result is defined as the percentage of positive results that are true positives for a defined population. The predictive value of a negative test result is the percentage of negative results that are true negatives for the same defined population. The predictive value of test results is directly dependent on prevalence of the condition in the population being tested (*37*). This is an important fact to remember when selecting tests.

The efficiency of a test is determined on the basis of the two predictive values. It is the percentage of all results that are true results, i.e., the number of true positives plus the number of true negatives divided by the total number of test results, expressed as a percentage (*37*).

Peak test sensitivity is desirable when a disease is serious and its diagnosis should not be missed, the disease is treatable, and false-positive results do not lead to serious problems. Peak test specificity is desirable when a disease is serious but is not treatable, the knowledge that the disease is absent has psychological or public health value, and false-positive results can lead to serious problems. A high predictive value of a positive test result is desirable when treatment of a false positive might have serious consequences. Peak test efficiency is desirable when the disease is serious but treatable, and when false-positive and false-negative results are about equally hazardous (*37*).

C. COLLECTION OF SPECIMENS

Concern for the quality of test results begins when a particular test is requested. There must be a system for the orderly and efficient requesting of tests; collection and identification of specimens; and transporting, preparation, and storage of specimens. Nothing is more important than having an adequate amount of an appropriate specimen in good condition for examination. If each specimen is not properly collected, labeled, and handled, or is not representative, the laboratory may do more harm than good by testing it (*22*). Some method is also needed for monitoring this total system (*121*). In many laboratories this latter process involves retrospective examination of records, but in other cases more sophisticated techniques are used (*57*).

D. CONTROL AND REFERENCE SERA

Control sera must be stable and capable of being stored for long periods without loss of activity. They must be similar in composition to patient specimens and must be subjected to the entire procedure being monitored. If sera are diluted or treated before being used as controls, their reactions must be verified to be similar to those of patients' sera. Controls should not be pretreated or diluted in bulk if such steps are included in the test procedure.

1. Source

Some control sera are available commercially. Small volumes are generally available as components in kits but are intended to be used only with a single kit. A few may be available in larger quantities.

Another way to obtain control sera is to pool patient sera. For many tests this source should not be overlooked, particularly if many samples with the desired concentration or reactivity levels are encountered during testing. However, the fact that patient samples may react with each other when pooled limits their usefulness.

Larger amounts of serum for serologic testing can sometimes be obtained from convalescent patients who have high titers and who are physically able to donate.

2. Preparation

Sera to be used as controls should be kept sterile to avoid deterioration. Sera with high concentrations or reactivity levels may be diluted with sera with lower concentrations or reactivity levels to obtain the desired levels. The normal level of the constituent to be tested should be considered in selecting the levels of the control sera. In general each procedure should have a normal control serum (negative), a strong positive control serum and another positive control serum which is reactive at the critical concentration (borderline positive). With some tests, controls with a low concentration of analyte should be included. Controls recommended by the manufacturer of a particular test should always be used and additional control sera can be included if a test involves special problems.

The serum should be assayed repeatedly in parallel with a serum of known concentration to ensure that the new control gives consistent results. The adequacy of the new control should also be checked by testing it with other procedures. The concentration or titer of the new control can also be verified by other laboratories. The serum may be filtered through sterile filters or preserved with 0.1% sodium azide or thimerosal (Merthiolate).

3. Storage

Enough control serum should be available to last for at least 6 months of continuous monitoring of test performance. Adequate storage must be provided to prevent deterioration.

Sera for most tests can be frozen and held at –20°C for extended periods, but should not be thawed and refrozen. Because of this fact, refrigerators with automatic defrosting should not be used. Serum can be dispensed in small aliquots (sufficient for one run), sealed tightly, and frozen. Many sera can be lyophilized and stored at refrigerator temperatures if facilities are available.

4. Reference Materials

Sera to be used as standards should be standardized against international reference materials when they are available (see Section IV). "Standards" included in commercial kits are not calibrated with each other and often are not interchangeable.

E. REAGENTS

Quality reagents are necessary for quality performance. A record should be kept of any changes in reagents in case the performance of a test changes. Before new reagents are introduced into a system they should be tested in parallel with the old reagents against a panel of appropriate reference sera to be sure that consistent reactions are obtained. The results obtained with the panel should reflect the sensitivity and specificity of the reagents being compared. Reagents should be stored according to manufacturer's instructions. Improper storage is a frequent cause of loss of reagent activity (*31*). Expiration dates must be observed. The pH's of buffers should be checked before each use. Each time reagents are used they should be examined for precipitates, cloudiness, contamination, or other signs of deterioration and should be discarded when there is evidence of their inadequacy.

Reagents should be labeled clearly to indicate their identity, hazards involved in their use, recommended storage conditions, and preparation and expiration dates (*19*). In some cases, such information as the dilution of stock reagents to be used in the test may also be helpful. Always follow the manufacturer's or author's directions in preparing and using the reagents.

Some manufacturers of reagents for serologic tests voluntarily submit samples of their products to CDC for the Premarket Evaluation Program which includes evaluation of reagents for rubella HI, bacterial antigens and control sera, viral antigens and antisera, and syphilis serology reagents. The reagents are tested according to CDC specifications using CDC reference reagents. On the basis of the reports they subsequently receive, laboratorians can request that manufacturers supply reagents from lots which have been confirmed as meeting CDC specifications.

Lists of some commercial sources for reagents are available (*31*). Because of the rapid changes occurring in the field of immunology, it is difficult to keep such a list completely up to date, but the Biological Products Evaluation Branch at CDC periodically issues such a list.

F. EQUIPMENT AND INSTRUMENTS

All glassware used in immunologic tests must be clean and free of detergent. Chipped or etched glassware should be discarded. Calibrated glassware should be checked for accuracy.

The user's accuracy and precision requirements should be met or exceeded when equipment is tested under working conditions. The manufacturer's specifications for performance should be checked and met. Instruments and equipment should be monitored routinely. The temperatures of water baths, incubators, refrigerators, and freezers should be checked periodically and records maintained. Maintenance should be performed and records kept on a regular basis by individuals who are trained and are familiar with the equipment.

Instruments used for measurements including spectrophotometers, spectrometers, dilutors, and automatic pipettors should be calibrated on a regular basis.

The monitoring necessary for commonly used laboratory equipment is listed in Tables 1 and 2. Recording thermometers are useful, as are centralized alarm systems for refrigerators and freezers. If such equipment is not available, other provisions should be made for routine monitoring (*31*).

Table 1. Suggested Monitoring of the Temperature of Laboratory Equipment

Item	Record Temperature	Other
Autoclave	Each run (also time recorded)	a. Record pressure once during each run. b. Use properly placed color indicator discs or strips in each run. c. Use peak temperature thermometer weekly. d. Use spore strips or spore suspensions monthly. e. If evidence of contamination is found, make sample cultures frequently (daily or weekly) until the cause is determined and eliminated.
Incubators	Continuously, with a recording thermometer (if possible); an alarm system is recommended.	a. If a recording thermometer is not used, record the temperature (a.m.) daily and before opening.
Water Baths	Daily before use.	a. Clean monthly.
Refrigerators	Daily; 2° to 8°C	a. Walk-in should have recording apparatus. b. Connect walk-in to alarm system. c. Clean monthly. d. Defrost or check refrigerator and freezer compartment every 3 months.
Hot Air Oven	Each run (also time recorded)	
Freezers	Daily	a. Connect to alarm system. b. Clean every 6 months.
Room Temperature	Each run (if applicable)	

Table 2. Suggested Monitoring of Laboratory Equipment

Item	Monitor Procedure
Analytical Balance	a. Check with certified weights at least once each week of use. b. Clean balances and weights monthly. c. Discard damaged weights which are no longer accurate.
pH Meter	a. Compensate for temperature each run. b. Date buffer solutions when first opened and, if possible, check monthly with another pH meter. Discard buffer solution if the pH deviates more than ± 0.4 from the manufacturer's stated pH or if it is contaminated with microorganisms. c. Standardize with at least one standard buffer (e.g., pH 7.0, if working in range below pH 6.0 use pH 4.0) before each test or series of tests. d. Inspect every 6 months for proper function.
Spectrophotometer	a. Check transmittance each day of use at a specified wavelength. b. Inspect every 6 months for proper function.
Centrifuge	a. Check brushes and bearings every 6 months. b. Check rheostat control against a tachometer at various loadings and frequently enough to assure proper gravitational fields. c. Performance must be evaluated often enough to assure proper performance. d. Oil and grease. e. Examine high speed heads for damage or corrosion.
Microscope	a. Clean after each use.
Filters	a. Do not sterilize filters (Millipore, etc.) above a temperature of 121°C. b. Carefully perform bubble point testing during each filtration run under positive pressure.

Table 2. Suggested Monitoring of Laboratory Equipment - Continued

Item	Monitor Procedure
Sterile Air Cabinets	a. Clean filters every 3 months.
	b. Check for leaks and rate of air flow every 3 to 6 months by measuring the pressure on both sides of filters.
	c. Expose blood agar plates to the air flow for 1 hour every month.
	d. Clean ultraviolet lamps every 2 weeks by wiping with a moistened cloth.
	e. Test ultraviolet lamps every 6 months, and replace lamps if they emit less than 70% of their rated initial output.
Water Stills	a. Connect distillation phase to the deionized water system.
	b. Drain and clean monthly according to the manufacturer's instructions.
	c. Clean distilled water reservior concurrently with the stills.
	d. Check distilled water continuously against suitable standards.
	e. Check pH conductivity or resistance daily to detect a leaking condenser coil.
	f. Check deionized water cartridges at regular intervals.
	g. Know what chemicals are put in steam lines and the frequency and concentration used.
	h. Check distilled water daily for resistance which should read a minimum of 0.5 milliohms.

G. PERFORMANCE OF TESTS

Performance is monitored with controls. Correct performance of reagents is reflected by the proper reaction in tubes which lack one or more of the components necessary for the reaction. For example, the presence of active streptolysin O reagent is demonstrated by hemolysis in the tube containing the reagent buffer and cells but no antibody to inhibit hemolysis.

Control sera provide a means of demonstrating the overall quality of performance of a test that detects or measures antibody at several concentrations or levels of reactivity. In most cases, a negative or normal level serum should be included to rule out nonspecific reactions. When inhibitors are known to be prevalent, as in rubella hemagglutination-inhibition, the negative serum should contain inhibitors to show the effectiveness of the treatment used to remove them. A serum with a titer at the level which is critical for interpretation (the borderline of normal and abnormal) should be used to monitor the sensitivity of the test.

H. REPORTING AND RECORD KEEPING

Complete and accurate records must be maintained in a good quality control system. These records should include personnel information; details of equipment preventive maintenance, service, and repair; copies of reports to physicians or other clients; accession records; records of reagents and materials used; records of observations made concurrently with the performance of each step in the examination of specimens; proficiency testing results; and internal quality control results.

The accessioning and reporting system should minimize the possibility of clerical errors. Precautions should be taken to prevent reporting results on the wrong specimen and transposing digits in reporting quantitative data. The system should be so designed that the history associated with a sample can be reconstructed in detail if necessary. Who performed which tests, what reagents and lot numbers they used, what the control results were for that run, and how and when the results were reported should also be documented.

To assist the physician interpreting laboratory reports and to increase the likelihood that the results will be used properly, the laboratory report should indicate the significance of the results by providing data on the values obtained by their methods on normal individuals and individuals with various abnormal conditions in the local population with the methods used in the reporting laboratory. With so many available tests, kits, methods, and commercial products and their differing sensitivities, reporting units, and significance levels, it is very difficult for any one physician to remember all of these variables when interpreting test results.

I. IMPLEMENTATION OF QUALITY CONTROL

It has been recommended that quality control systems be established in gradual steps so that the implementation process can also be an educational process and so that the sudden introduction of all the available techniques does not overwhelm the laboratory staff (*121*).

The recommended stages are as follows:

1. Optimal conditions variance.
2. Routine conditions variance (known values).
3. Routine conditions variance (unknown values).
4. Statistical use of patients' results.
5. Interlaboratory comparisons.

Stage 1 involves evaluating the test to determine the variation that exists under ideal conditions. At least 30 tests should be run and the results plotted. These data provide the basis for preparing a quality control chart (*56*). If the results are normally distributed, the usual parametric statistical methods can be used (*100,101,121*), but if the literature or results of the K-S test (see Section VI) indicate that results are not normally distributed, a transformation must be performed or other calculation methods used. Immunologic results are frequently lognormally distributed, and the calculation methods described in Section VI can be used to prepare quality control charts.

In stage 2 the variance is measured on known samples under routine conditions. Variance from these results can be used to establish control charts which will probably have wider limits but can be the basis of day-to-day quality control. Stage 3 is similar to stage 2 except that the samples are submitted as unknowns to eliminate bias in reading the test results.

Quality control in stage 4 is based on the observation that the mean patient value of any analyte tends to remain constant if a sufficiently large sample is measured and if obviously high or low values are excluded. Monitoring mean patient values provides additional verification that other quality control processes are working and also allows the detection of changes which cannot be found with other quality control methods (*57,121*). Analysis of patient data can reveal errors in sample collection and handling and clerical errors not otherwise detectable.

The final stage consists of participation in interlaboratory comparison surveys; the results provide data on how the precision and accuracy in one laboratory compare with those in another and can lead to measures facilitating uniformity and standardization. Every laboratory should participate in some proficiency testing program and every person in the laboratory should participate in some intralaboratory performance evaluation activity (*29*). Proficiency testing programs are designed to provide information on the relative performance levels of laboratories and those of various technologists. If proficiency testing samples are processed in the same way as patient samples, the resulting information is useful in determining training needs, inferior methods, weaknesses in intralaboratory quality control, and in evaluating an individual technologist's performance. Such programs should be part of the total program of professional development and training of laboratory personnel.

III. QUALITY CONTROL OF SPECIFIC PROCEDURES

A. PRECIPITIN TESTS

1. Principles Involved

Antigen and antibody react in solution by precipitating to form detectable aggregates. The primary reaction of antigen and antibody is followed by secondary reactions which produce the visible result. Specific crosslinking of antigen-antibody complexes results in aggregates which continue to grow in size until they fall out of solution. Antigen-antibody interactions may mask individual polar groups which decrease the solubility of the complexes and increase aggregation. The reaction peaks when the antigen and antibody are in optimal proportions and occurs most rapidly then, and neither antigen nor antibody are present in the supernatant. The rate of precipitin formation also depends on the type of antibody, temperature, salt concentration, pH of the suspension, and the lipid and protein constituents of the serum.

Several types of precipitin reactions are used in clinical serologic tests. Antigen and antibody can react both in solution and in a gel matrix (immunodiffusion). Antigen and antibody solutions are mixed in capillary tubes and observed for precipitation as in tests for C-reactive protein or those for grouping and typing streptococci. Precipitation in dilute solutions can be detected with nephelometry. Various types of instruments have been developed to detect the light scattered as a result of the formation of aggregates, e.g., fluorometers, nephelometers (with tungsten lamps or lasers), and rate analyzers. Adding polyethylene glycol to the solution may heighten the antigen-antibody reaction. Nephelometric procedures are readily automated.

Immunodiffusion methods also involve precipitin reactions. In double immunodiffusion, the reactants diffuse through a gel and precipitate to form a line at the point where their relative concentrations are optimal for precipitation.

The position and curvature of the lines formed are related to the concentrations of the reactants and their diffusion coefficients. In a properly controlled test, the number of bands indicates the minimal number of antigen-antibody systems present. Antigens (or antibodies) may be compared by double diffusion when wells are spaced so the precipitin lines meet. The lines will fuse if the systems are identical, they cross if they are not the same, and spurs form if there is partial identity. Double diffusion is used frequently in fungal serology.

Rheophoresis (*118,119*) is a modification of double immunodiffusion in which the diffusion of the reactants is influenced by the movement of the buffer. The water in the gel is allowed to evaporate through an opening in the cover over the area where the reaction is to occur. Buffer is placed at the edge of the plate and as it flows through the gel to replace the water being lost by evaporation, the reactants are brought together. Another modification of double immunodiffusion, counterelectrophoresis (*39*), involves using electric current to bring the reactants together (discussed under Electrophoresis in part H of this section).

Single radial immunodiffusion is frequently used for quantitating various serum proteins, primarily the immunoglobulins. With this method, antibody is incorporated into the gel, and the antigen is allowed to diffuse from a well cut in the gel. As the antigen diffuses, it is precipitated by the antibody, forming a ring or disc around the well. The ring enlarges until the reactants are in equilibrium. The size of the ring is related to the amount of antigen present in the well and the original concentration can be calculated with proper standards.

2. Problems Frequently Encountered

Precipitin tests are generally less sensitive and the reading of them more subjective than other tests. Gel diffusion methods and instrumentation have helped to overcome some of these difficulties and make precipitation a practical quantitative method.

Antigen-antibody proportions are still a problem because the reaction is not detectable in the presence of a large excess of either reactant. Some provision to distinguish such excesses from the absence of one reactant must be made.

3. Control Recommendations

Standardized reagents should be used for precipitin tests when possible. Standards (calibrators) should be calibrated against available international reference preparations. Standards and controls supplied by different manufacturers often do not show comparable results and thus cannot be used interchangeably. Standards should be included in each run, and unknowns with concentrations outside the range of the standard sera should be rerun after dilution if they are too high or should be tested in another system. Control sera should also be used as a continuous check on the system, and results should fall within the preset limits (see Section VI).

The methods of preparing a working standard serum calibrated against the WHO standard for immunoglobulins and of high and low internal standard sera and their use in quality control have been described elsewhere (*79*).

For single radial immunodiffusion, measuring the amount of antigen may lead to better precision than just filling the wells and risking overfilling or underfilling. Wells should not contain fluid before antigen is added to avoid dilution. The temperature during incubation should be kept relatively constant, the incubation time should be consistent, and reference sera and unknowns must be read concurrently. Plates should be incubated on a level surface so that the rings obtained are symmetrical. Nephelometric procedures require clear sera and reagents to minimize nonspecific light scatter. A blank may be required for each serum, depending on the method.

When the prozone phenomenon is occasionally observed in the tube precipitin test, the antigen rather than the antiserum should be diluted because when excessive antigen is present, soluble complexes of antigen-antibody are formed rather than the insoluble complexes which form the precipitate. Prozone reactions rarely occur in the agar gel tests because at some point during migration of the antigen and antiserum through the agar, the concentration of reagents is such that an insoluble precipitate is formed (*45*).

4. Standardization

Commercial reagents for immunoglobulin quantitation are fairly well standardized against the WHO reference preparation. Reagents for other serum proteins are not standardized, and conflicting results may be obtained with different reagents. Some work is being done to provide standardized reference materials for other constituents.

B. AGGLUTINATION TESTS

1. Principles Involved

Agglutination is the aggregation of insoluble particles which may range in size from microscopic bentonite particles or bacteria to red blood cells (RBC). Such aggregates are usually macroscopic. Agglutination is used in serologic tests in several forms. Direct agglutination involves the reaction between an antibody and a particulate antigen (*118*). This reaction is used in such procedures as bacterial agglutination tests and heterophile agglutination tests for infectious mononucleosis (IM).

The number of antigens that can be used and the sensitivity of the agglutination test may be extended by attaching soluble antigens to the surface of serologically inert particles such a bentonite, latex, or RBC. Once these particles are coated, their role in the test is essentially passive (*109*). Antibody reacts with the antigen and the particles agglutinate, but only because of the reactivity of the attached antigen. Antibody may also attach to particles which are then used to detect the presence of antigen. When antigen is attached to the particle, the test is passive (or indirect) agglutination; if antibody is attached, it is reverse passive agglutination. Passive agglutination is used in tests as diverse as those for rheumatoid factor (RF), toxoplasmosis, and thyroid auto-antibodies. Reverse passive agglutination is frequently used for hepatitis Bs antigen tests.

Agglutination tests are more sensitive and rapid than precipitation tests. In general, IgM antibodies are more efficient in agglutination than are IgG antibodies, because of their size as well as their additional binding sites. Occasionally IgG antibodies are unable to cross-link sufficiently to cause agglutination and have been termed "incomplete" or "blocking" antibodies. Additional measures may be necessary to facilitate agglutination when they are present. Depending on the particular antigen-antibody system, centrifugation, enzymatic treatment of RBC, or changes in pH or ionic strength may enhance the reaction.

2. Problems Frequently Encountered

Sensitivity of agglutination tests may vary considerably with the antigen used. In bacterial agglutination tests, geometric means of proficiency testing (PT) results show definite differences among commercial products. The density of the organisms and the strain used to prepare the antigens affect the sensitivity. The age of sheep red blood cells (SRBC) used in heterophile agglutination tests for IM may influence the results. Cells from different animals may respond differently in the test system. Commercial reagents for rheumatoid factor (RF) tests vary widely in sensitivity.

Problems affecting the specificity of agglutination tests include cross-reactions, auto-agglutination of the antigen, and interfering antibodies. Some bacteria share common antigens and therefore react with heterologous antisera. For example, some *Brucella* antisera will also react with *F. tularensis*, which may produce misleading results. Cross-reactions between some rickettsia and *Proteus* strains are the basis for the Weil-Felix tests. Auto-agglutination of the antigen can sometimes be a problem when the antigen suspension is not stable. Antibodies other than those expected in the serum may pose a problem — particularly with passive agglutination tests. When coated RBC are used, antibodies to red blood cell antigens may still cause agglutination. RF in sera being tested in reverse passive agglutination tests may react with the antibodies (IgG) which are coated on the particles.

Antisera are occasionally found that do not agglutinate antigens at lower dilutions but do so at higher antibody dilutions. This selectivity is referred to as the prozone phenomenon and results from the high concentration of antibodies at the lower serum dilutions which is reduced as the serum is diluted. In some antisera an apparent prozone phenomenon may be caused by the presence of nonagglutinating antibodies (*45*).

3. Control Recommendations

To obtain reproducible results in agglutination tests, technique must be consistent. The concentration of antigen, the diluent, and time and temperature of incubation must always be the same for a particular test. A simple way of estimating the concentration of a particulate antigen is comparing it against a McFarland Standard (*61*) or against opacity reference preparations. When soluble antigens are used to coat RBC's for the indirect hemagglutination test or for coating inert particles like latex or bentonite, it may be necessary to use a chemical or photometric method for determining the amount of antigen present (*45*).

Available standardized antigens and reference materials are helpful in obtaining consistent results in agglutination tests. Positive and negative control sera are required to monitor performance of these tests. The positive sera, one high titered and one borderline, are used to check the sensitivity of the test. Results should be within the limits previously established for the sera (see Section VI). The negative control serum, which should not react with the antigen, provides a check on the specificity of the test. A control tube in which saline or buffer replaces the serum must also be checked to confirm that the antigen does not agglutinate nonspecifically (auto-agglutination). For passive agglutination tests, the uncoated particle should not react with the serum being tested.

Paired sera, i.e. acute and convalescent sera, should be tested concurrently to minimize the effects of any differences in reagents or conditions.

Tubes should be examined for agglutination in a prescribed manner according to the standard technique. Bacterial or latex agglutination tests should be read against a dull black background with a strong light shining through the tubes. A white background is better for reading RBC agglutination. The end-point reaction (least reaction regarded as positive) should be consistent.

Controlling slide agglutination tests involves additional problems. Only a few tests should be run at a time, so the mixtures will not dry out and produce confusing results. Thorough mixing is important, and the slide should be rotated according to instructions. The proper light and background are also necessary for reading reactions.

4. Standardization

Interest has been shown in standardizing at least two agglutination tests, those for *Brucella* antibodies and RF. Standard antigens are available for *Brucella*, and commercial companies have been encouraged to standardize their antigens to be comparable to the National Animal Disease Laboratory antigen. Reference antibody has been prepared for the World Health Organization (WHO). At present, however, neither the standard antigen nor the reference antibody is widely used. Manufacturers, Government agencies, and professional organizations have shown more interest in standardizing the RF tests. WHO has a reference preparation for RF, and CDC and the National Committee for Clinical Laboratory Standards (NCCLS) have prepared another. The use of such preparations has been shown to decrease the variability of results in proficiency testing surveys (*36,109*).

C. COMPLEMENT FIXATION TESTS

1. Principles Involved

The complement fixation test is based on the principle that when a certain antibody combines with its homologous antigen, complement is fixed and therefore is not available to lyse the sensitized cells of the indicator system. The test consists of two separate antigen-antibody reactions. In the first step the antigen and the antiserum are combined. If the antiserum contains an antibody which is specific for the antigen, an antigen-antibody complex forms which inactivates or "fixes" complement. In the second step a suspension of red cells which have been sensitized by the attachment of antibodies (hemolysin or amboceptor) is added. If complement is available (indicating that it was not fixed in the first step), it reacts with the sensitized cells to produce hemolysis. If the complement is fixed in the first step, no reaction takes place in the second step. Therefore, hemolysis indicates that an antigen-antibody reaction did not take place in the first stage, and no hemolysis indicates that an antigen-antibody reaction did take place.

For the test to proceed properly, all components of the system except the unknown must be balanced. The amounts of complement and hemolysin must be titrated to assure that there are levels that will permit hemolysis if complement is available (an antigen-antibody reaction does not take place) and that no hemolysis will take place if complement is not available (an antigen-antibody reaction does occur). The antigen and the antiserum must also be individually tested to verify that they do not possess "anticomplementary" activity, i.e., that one does not fix complement without the other.

Other variables which must be considered in performing this test are: pH, ionic strength, osmolarity, Ca^{++} and Mg^{++} concentration of the diluent buffer, temperature, test volume, quality of sheep erythrocytes, and cell preservatives. Significant changes in any of these can alter the test results.

Because of the number of variables involved in the test, acute and convalescent phase sera should be tested in the same test run to determine whether there has been a change in titer.

2. Problems Frequently Encountered

Most problems with this test concern the delicate balance of reagents and conditions necessary. Of special importance is the proper complement dilution used in test.

Poorly reactive antigens and anticomplementary serum or antigens are frequent problems. Serum can become anticomplementary as a result of bacterial contamination introduced by improper collection or handling. This can be prevented by using aseptic technique or by adding thimerosal (Merthiolate) to the serum. Thimerosal concentrations of 1:10,000 or 1:5,000 retard the development of any anticomplementary activity without significantly affecting the complement, but higher concentrations should not be used because they may inactivate the complement (*117*).

3. Control Recommendations

Minimum Controls

a. Hemolysis control (buffer control) contains sensitized cells and complement, but no antigen or antibody. The purpose of this control is to indicate that sufficient hemolysin was used to sensitize the red blood cells and that the cells will be lysed in the presence of complement.

b. Antigen controls are included to assure that the antigen will react with specific antibody and is not anticomplementary. This is done by testing the antigen in dilutions of known positive and negative sera.

c. Serum controls determine that serum is not anticomplementary at any dilution used, by not adding antigen to these tubes.

d. The RBC control is included to show that the cells do not lyse spontaneously in the absence of complement.

e. Tissue controls are used with viral antigens and are uninfected tissue of the kind used to prepare the antigens. They are included to show that the uninfected tissue was not the reason for the fixation of complement.

Additional controls which are sometimes used include several different concentrations of complement in the antigen and antiserum controls (for example, 5, 2.5, and 1.25 $C'H_{50}$ units); known positive and negative antiserum controls; and known positive antigen controls (*45*).

It is recommended that complement be titrated and that the complement titration curve (% lysed/% nonlysed cells vs. dilution of complement) be plotted on logarithmic (log-log) paper. The slope (1/N), from van Krogh's formula, should be calculated and should have a value of 0.20 ± 0.02 (*117*).

Every new lot of reagent should be tested in parallel with one of acceptable reactivity before it is used in routine testing.

4. Standardization

Various CF tests developed over the years have primarily been modifications introduced by individual investigators studying particular organisms. The major differences among them involve the concentration and volume of test reagents and the time and temperature of incubations. The Laboratory Branch CF (LBCF) test (*117*) was developed at CDC to standardize the test and eliminate the need for different procedures for each antigen. A microadaptation of this standardized test has also been reported (*74, 117*). The (LBCF) test is suitable for most antigen-antibody systems for which the CF test is usable.

"Checkerboard" or "block" titrations provide a means of improving inter- and intra-laboratory precision and play a major role in the standardization of diagnostic antigens.

D. INHIBITION TESTS

1. Principles Involved

Many serologic inhibition tests assay for the presence of antibodies which prevent or inhibit the occurrence of certain characteristic reactions. Typical reactions in these protocols include hemagglutination, as in certain virus antibody assays, and hemolysis during antistreptolysin O (ASO) determinations. The hemagglutination phenomenon is a direct agglutination and involves an aggregation reaction between virus particles (antigen) and RBC. In a serum containing homologous antibodies, the virus is bound and thus unable to agglutinate the RBC, i.e., hemagglutination is inhibited.

The test for detecting ASO is based on the fact that streptolysin O (SLO) reacts enzymatically with RBC to effect their lysis. ASO specifically neutralizes SLO, inhibiting its hemolytic activity, and the amount of ASO may be quantitated by using a series of standard dilutions (*76*). Antibodies to other bacterial exoenzymes may also be detected with inhibition assays.

Both examples of the inhibition described above are basically results of the neutralization of the antigens by specific antibodies, which results in the formation of soluble Ag-Ab complexes.

2. Problems Frequently Encountered

Sera frequently contain nonspecific inhibitors (NSI) of viral hemagglutination which must be removed before the inhibition test can be performed. The presence of NSI results in a false-positive reaction. Standardized procedures for removing NSI from sera that are to be screened for rubella HI antibodies have been described (*24*). Occasionally, the methods currently used for removing NSI from sera are unsatisfactory (*92,95*). Aging of sera may trigger structural changes in the β-lipoprotein fractions of certain sera which in turn may prevent the usual precipitating complexes from forming between the NSI and the serum-treatment reagents. The kind of treatment being used may make a difference in how well NSI are removed (*82*); also final treatment of the serum with Na_2CO_3 can reduce the concentration of divalent cations which can suppress NSI expression (*13*).

Human sera may contain other nonspecific agglutinins besides NSI for the RBC used in the test. These too, must be sorbed out before valid test results can be obtained.

When selecting an antigen for the HI test, one should be sure that its potency (HA titer) is adequate to attain the sensitivity required. Rubella antigen preparations with low titers (less than 64), may contain nonhemagglutinating viral particles which compete with hemagglutinating particles for antibody. The results of using a low-titered antigen could be lower antibody titers or false-negative titers. Therefore, each lot or vial of antigen should be titrated before use. Using only antigens with titers equal to or greater than 64 will reduce the effect of nonhemagglutinating particles.

Other factors which influence the sensitivity of the HI test are the kind and age of the RBC used. For rubella HI, trypsinized human O cells are the most sensitive and tend to produce titers with higher geometric means than do goose, baby chick, or commercially prepared chick RBC (*110,111*). Incubation temperatures and the pH of the diluent may also affect test results. Because HI tests are complex procedures and decisions about patient care may be based on their results, they should be performed by well-trained, experienced individuals.

One of the main difficulties associated with the ASO test is maintaining SLO in a reduced state. In the conventional ASO procedure, SLO plays two parts. It is the antigen which combines with ASO, and it serves as part of the indicator system of the test by lysing the RBC when ASO is absent. Lysis by SLO can occur only when it is in the reduced state. Oxidation prevents SLO from performing that function. SLO obtained commercially is usually a lyophilized preparation. It should be reconstituted with water that has been cooled to 4°C and kept at that temperature until it is used in the test. Water temperatures above 4°C promote oxidation of the SLO and reduce its potency, thus causing falsely high ASO titers.

False-positive ASO titers can be caused by increased levels of β-lipoproteins in a patient's serum, but these compounds may be precipitated out with dextran sulfate-calcium chloride. Certain strains of bacteria (*B. cereus, Pseudomonas* sp.) may also cause false-positive ASO titers as contaminants of the test serum.

3. Control Recommendations

Because of the many factors in the rubella HI test which can adversely affect results, controls used should be able to monitor each of the possibilities very closely. Antigen concentration should be properly adjusted to the recommended number of hemagglutinating units, and a confirmatory back titration should be performed during the test. The sensitivity of the antigen and the laboratorians's skill in making dilutions should be checked by using sera of three different HI titers: one high (> 128), one low (16-32), and one negative (< 8). These control sera and the test sera should be treated concurrently to remove NSI and nonspecific agglutinins (NSA). The HI titers should then fall within the limits previously established. Other controls should include a serum control to monitor removal of nonspecific agglutinins from the serum and cell control to document the absence of auto-agglutination of the RBC.

The controls used for monitoring the reliability of ASO determinations are similar to those already described. These include a cell control to show that the RBC do not lyse spontaneously, a known positive ASO control, an SLO control to show the RBC are lysed by the reagent, and serum controls. Some sera will contain natural hemolysins for the indicator RBC which will be detected by lysis in the serum controls (serum plus RBC). However, the presence of hemolysins will not invalidate the test results as long as they have been removed by dilution before the end point has been reached and if there is at least one dilution that shows no hemolysis.

4. Standardization

It is often difficult to standardize a test protocol when many variables are involved. However, as the number of variables increases, so does the need for standardization. Laboratories are therefore advised to use a standardized method. A laboratory can attain 93% or better precision if it follows the protocol as outlined in the "CDC Standardized Hemagglutination-Inhibition Test" (*24*). Proficiency testing surveys show that most of the variation in rubella HI results is caused by some departure from the standardized procedure (*110,111*). (This procedure was designed to minimize variation, yet maintain the necessary sensitivity and specificity.) At one time, kaolin was used for extracting NSI from the serum; heparin-$MnCl_2$ and dextran sulfate-$CaCl_2$ NSI precipitation techniques are now preferred. However, of all the laboratories surveyed, 44%-55% continue to use the kaolin procedure (*110,111*).

The most commonly used ASO procedure is that described by Todd in 1932 (*113*) and modified by Rantz and Randall in 1945 (*83*). An ASO procedure based on the same general principles has been developed for use with the Microtiter system (*53*). The primary differences are the volumes of reagents used and the dilution schemes. A linear dilution series is used in the Rantz-Randall procedure and a logarithmic dilution series in the CDC method (*53,76*), which is simpler to prepare than the dilution series of Rantz-Randall. Most SLO reagents are standardized by the manufacturers.

E. IMMUNOFLUORESCENCE MICROSCOPY

1. Principles Involved

Fluorescent dyes, such as fluorescein and rhodamine, may be conjugated to antibodies with only minor changes in the reactivity of either. These dyes absorb light of short wavelengths and emit light of longer wavelengths. Generally fluorescein is preferred for immunofluorescence tests because its green fluorescence is easier to see than the red fluorescence of rhodamine and because red autofluorescence is more common than green (*67*).

Fluorescein isothiocyanate (FITC) is the form most commonly used for clinical work. It is excited at wavelengths between 350 and 495 nm (maximum at 490-495 nm) and emits light at 520-530 nm. The optical system capitalizes on the difference in these wavelengths. Light is provided by a mercury vapor lamp (xenon or quartz-halogen can also be used). The light passes through the excitation filter which selects light of the proper wavelength for excitation. A darkfield condenser is usually used so that only fluorescence is seen against a dark background. A barrier filter is used between the slide and the ocular to filter out all light except that of the emission wavelength.

Incidence illumination may also be used for fluorescence microscopy. Light is provided above the slide and specific wavelengths are reflected by dichroic mirrors. By using selective filters with dichroic mirrors, the system reflects light with wavelengths of 480-500 nm to the slide and allows light with wavelengths of 500-530 nm to pass from the slide to the ocular.

Immunofluorescence may be used in various assays. Direct assays are used to detect or localize antigen in tissues or to identify organisms fixed to a slide, and indirect assays are generally used to detect antibodies in a patients' serum. In the direct methods, the specific antibody is conjugated with FITC and allowed to react with the antigen on a slide. Positive reactions consist of fluorescence of the antigen-antibody complex. For indirect methods the patient's serum is allowed to react with the antigen on the slide, the excess or nonreacting proteins are washed off, and conjugated antihuman globulin is added. The conjugated antihuman gamma globulin reacts with the specific antibody which is attached to the antigen. A positive reaction is seen as fluorescence of the complex. The conjugated antibody may be species-specific (e.g., antihuman globulin) or it may be class-specific (e.g., human anti-IgM).

Immunofluorescence is used for a number of different antigen-antibody combinations. Antinuclear antibodies, other autoimmune antibodies, syphilis antibodies, and antibodies to toxoplasma and viral agents are routinely assayed with indirect immunofluorescence. Direct immunofluorescence is used to group streptococci and to detect antigens in histologic preparations or organisms in direct smears.

2. Problems Frequently Encountered

One of the problems encountered with immunofluorescence techniques is in differentiating between nonspecific staining and specific staining resulting from the antigen antibody reaction. Nonspecific staining is caused by the adherence of labeled proteins to the slide or tissue rather than being an immunologic reaction. Generally nonspecific staining is diffuse rather than discrete, but it depresses the background contrast and makes specific reactions more difficult to see. Conjugates with higher stain to protein ratios generally produce more nonspecific staining; however, if they have reasonably high titers and are used diluted 1:16 or greater the nonspecific staining may be low in conjugates with high stain to protein ratios. Allowing the slide to dry before each reagent is added may also increase nonspecific staining. Staining of specific soluble antigens is frequently encountered and can be diagnostic when proper controls are used. Specific staining of antigens other than the expected one may also be a problem, but one which can usually be remedied by adsorbing the antibody (*25*).

When FITC is exposed to light and is excited, some of its electrons are irreversibly dissociated, and the dye fades or is quenched (*67*). Therefore, slides should be read immediately or should be placed in the dark for one or two hours before reading. If not, the fading of fluorescence will make them difficult to read.

Specificity of the antibody and conjugate can be a problem with tests as sensitive as immunofluorescence. Reading the slides is subjective, and results depend on the knowledge and experience of the person performing the tests.

3. Control Recommendations

Control sera should be included in all immunofluorescence assays. For the indirect test, positive control sera should be reactive within the limits previously set (see Section VI), and the negative serum should not produce significant levels of fluorescence. An additional control of buffer (instead of patient's or control serum) and conjugated antibody should produce no nonspecific fluorescence. If unrelated antigens could be a problem, they should be included in the controls. Two specificity controls which could be included are: (1) a conjugated antiserum which should not react in the system, i.e. antiserum to a different antigen, and (2) testing for inhibition of the reaction by addition of homologous (but not of heterologous) antigen to the conjugated antiserum (*31*).

For the direct test it is important to use known homologous and heterologous antigens as tests of specificity.

Characterizing the conjugated antiserum is an important control for the test because the titer is dependent on the fluorescein to protein (F/P) ratio. The F/P ratio for human tissue antibody testing should be between 3 and 5, and the antibody to total protein ratio should be greater than 0.1 (*23*). For other tests the ratios may be different but usually have an optimal range. A balance between nonspecific staining and the intensity of specific staining is dependent on a complex relationship among protein, fluorescein, and antibody concentrations, and each batch of conjugate has to be tested for optimal conditions (*25*).

The optimal dilution of conjugate to be used in indirect immunofluorescence tests must be determined by block titration of conjugate and antigen (*23,67*) for each test system. It should produce a low level of nonspecific staining, reproducible titers for known positive sera, and no fluorescence with negative sera or buffer control (*23*). For direct immunofluorescence tests the maximum dilution of conjugate giving 4+ staining is usually the best.

To obtain maximum fluorescence, the microscope must be properly maintained and adjusted. The following checks, or those specified by the manufacturer, should be performed routinely at frequent intervals (*45*).

(a) Align and focus the light source on the back lens of the condenser, being certain that the reflected image of the light filament is nearly superimposed upon the direct image of the filament.

(b) Align the light leaving the darkfield condenser with the objective.

(c) Check the light source to be certain that it has not been clouded or darkened by etching or by deposits of mercury on the inner surface. (Keep a log of use for the mercury lamps and replace them before or at the life span as stated by the manufacturer.)

(d) Remove all excess oil from the substage condenser and objectives.

(e) Remove all dust, lint, or other particles from the entire optical system.

4. Standardization

The standardization of diagnostic immunofluorescence is a complex problem because diagnostic results are greatly influenced by such interacting factors as equipment, materials, and techniques for expressing and recording fluorescence. Furthermore, the characteristics of immunofluorescence reagents are determined by how they are manufactured and used. Using stable reference preparations of such reagents appears to be the only practical way to standardize laboratory test results (*2,26,87*). Several professional organizations are actively promoting this objective.

F. RADIOIMMUNOASSAY

1. Principles Involved

Since it was developed in the late 1950's by Berson and Yalow, radioimmunoassay has become increasingly valuable in the clinical laboratory (*10*). Radioimmunoassay can generally be defined as an analytical method based on the reaction between an antigen, a radioactive labeled antigen, and an antibody. In a radioimmunoassay the concentration of the antigen (standard or unknown) is inversely proportional to the concentration of radioactive labeled antigen-antibody complex formed. The specificity and to a large degree the sensitivity of the method are provided by the antibody. The radioactive labeled antigen provides a means for measuring the products of the reaction and thus contributes to the sensitivity of radioimmunoassay.

Radioimmunoassays are based either on the principles of equilibrium analysis or those of sequential saturation analysis

(*132,133*). In both equilibrium analysis and sequential saturation analysis, immunochemical reactions between very small amounts of reactants are involved. Thus, in both cases the reactions are influenced by the chemical and physical environment—that is, the pH, ionic strength, and temperature. Thus these parameters must be controlled. Because some additional quality control measures for radioimmunoassay are dictated by which of the two sets of principles are used, they must both be clearly understood.

Equilibrium analysis is the technique used when labeled and unlabeled antigen are reacted with the antibody until a mass equilibrium is attained. The reaction can be written:

$$Ag + {}^*Ag + Ab \rightleftarrows Ag\text{–}Ab + {}^*Ag\text{–}Ab \qquad \text{(I)}$$

where,

Ag = unlabeled antigen
$^*Ag$ = radioactive labeled antigen
Ab = antibody
$Ag\text{–}Ab$ = unlabeled antigen-antibody complex
$^*Ag\text{–}Ab$ = radioactive labeled antigen-antibody complex.

The requirements for a workable assay are: (1) the equilibrium must favor the formation of the products on the right side of equation I, (2) Ag and $^*Ag$ must have identical affinity for Ab, (3) Ab must be specific for Ag and $^*Ag$, and (4) the concentration of Ab must be less than the combined concentrations of Ag and $^*Ag$.

When Ag and $^*Ag$ have different affinity for the Ab, it is necessary to use sequential saturation analysis. The reactions for sequential saturation analysis are:

$$Ag + Ab \rightleftarrows Ag\text{–}Ab \qquad \text{(II)}$$

$$^*Ag + Ab_1 \rightarrow {}^*Ag\text{–}Ab_1 \qquad \text{(III)}$$

Unlabeled antigen is first reacted with excess antibody, Ab, to form an antigen-antibody complex. After equilibrium is reached, $^*Ag$ is added to react with the remaining unbound antibody, Ab_1, to form the complex $^*Ag\text{–}Ab_1$. Reaction III is run just long enough for $^*Ag\text{–}Ab_1$ to form, that is, a new equilibrium is not established.

This general discussion of the fundamentals of radioimmunoassay is intended only as an introduction. References are provided for a more extensive review (*11,12,30,43,99,120,130*).

2. Problems Frequently Encountered

Both within- and between-assay precision may present major problems for laboratorians, particularly to those who are inexperienced with radioimmunoassays. Since substances present in a concentration of 10^{-9} to 10^{-12} moles per liter are generally measured by radioimmunoassay, even slight errors in technique or variation from the assay protocol can produce errant results.

Poor within-assay precision can result from imprecise pipetting, insufficient mixing of reagents (or samples), nonuniform mixing and centrifugation of assay tubes, poor technique in separating the bound from the free fraction, or faulty scintillation counting. Poor between-assay precision is usually caused by changes in reagents (such as pH, ionic strength, standards, or antibody), or by not adhering to assay protocol.

Single discrepant results may also be obtained. Results which are inaccurately very high or very low usually result from an error in a single reagent addition and are usually very obvious and easily remedied. An inaccurate result which is caused by the presence of cross-reacting or interfering substances in the sample may be noticed only if the requesting physician notes that it is inconsistent with the patient's clinical status or if the result is obtained on a sample tested in an interlaboratory survey where the result is compared to those from other laboratories.

Many other problems such as shifts in the standard curve, all samples in a radioimmunoassay having inaccurately low or high values, inappropriate levels of radioactivity for nonspecific binding or maximum binding samples, and a flat standard curve are all usually related to an error in reagent composition or reagent addition. These problems as well as most of those previously mentioned can be prevented or at least detected with a well-designed quality control program.

3. Control Recommendations

As is true for any analytical technique, the quality control program for radioimmunoassays must focus both on error prevention and error detection. Most problems can be prevented with proper equipment maintenance and calibration, with proper preparation and storage of reagents, and with proper training and education of the laboratorians who perform the tests. Problems in these areas can easily be detected with a quality control program which includes monitoring and documentation of the standard curve parameters, the results for routine quality control samples, and the laboratory's performance in interlaboratory proficiency testing.

Various pieces of equipment are used in the radioimmunoassay laboratory. Freezers and refrigerators used for storing specimens and reagents should have their temperatures recorded daily. Constant water bath and room temperature should be recorded when they could affect an assay, because a change from the required temperature would alter the reaction kinetics. The accuracy and precision of pipetting devices should be verified monthly, weekly, or even daily if they are extensively used. Balances and pH meters used in reagent preparation should be routinely scheduled for preventive maintenance and calibration. Centrifuges, used for separating bound from free antigen in some procedures must be calibrated routinely to assure reproducibility. The proper functioning of scintillation counters must also be confirmed. Background counts and counting efficiency should be recorded daily. Changes in line voltage or increases in radiofrequency waves generated by other equipment operating in the proximal area must be documented not to alter the performance of the scintillation counter.

All reagents used in radioimmunoassays must be carefully prepared and stored. Written protocols should be readily available for preparing buffers, for diluting antisera, and for preparing standards and other reagents. Protocols must include the source and required purity of reagents, pH and ionic strength, and the expected shelf-life. The sample to be analyzed can also be considered a reagent. The proper method of specimen collection and storage must be documented in detail.

Most radioimmunoassays are not fully automated; thus, trained laboratorians must perform the manual operations meticulously. Coincidental to this training, instructions about the theory of radioimmunoassay should be given. For any particular method, the laboratorian should be aware of when timing between steps in a procedure is critical, when the order of reagent addition is critical, and when small pipetting errors can cause erroneous results. Procedure manuals must give detailed instructions for pipetting, mixing solutions, centrifuging reaction tubes, and an account of the technique for separating the bound from the free antigen fraction. Laboratorians must also be taught methods for plotting standard curves and the conditions under which various plotting techniques can be applied.

Even with the aforementioned preventive measures, errors can and will occur. Thus a detection system for errors must be implemented. Control charts should be maintained on the standard curve parameters as well as on results obtained with quality control samples. Standard curve parameters which have been found useful are the dose intercept at 50% maximum binding, the percentage of blank (nonspecific binding) counts relative to total counts, the percentage of maximum binding (zero dose) counts relative to total counts, and the slope and y-intercept of linear standard curves. Discrepant values for these parameters are invaluable not only in detecting but also in troubleshooting errors. A shift in the intercept at 50% of the maximum binding may indicate improper antibody dilution or alteration in the antibody itself. An increase in the percentage of blank counts relative to total counts may signal insufficient separating reagent used. An increase or decrease in the percentage of maximum binding relative to the total counts can indicate an error in the amount of radioactively labeled antigen used. All of these pieces of information are readily available at the end of an assay.

The aspect of quality control with which laboratorians are most familiar is testing quality control samples. For a radioimmunoassay, in which the concentration of highest standard analyzed may be 100 times that of the lowest, at least three levels of quality control samples should be incorporated into each analytical run. In radioimmunoassays based on the sequential saturation technique, a set of quality control samples should be placed at both the beginning and end of runs which include a large number of tubes to assure that the time differential created by pipetting, vortexing, and other individual tube manipulations does not cause samples at the beginning of the run to react differently from those at the end.

As stated earlier in this monograph, quality control samples must react comparably to patient samples in the test system. Laboratories which cannot prepare their own control sera have access to commercial materials for most analytes. However, the laboratory must verify parallelism between the expected control values and the standard curve by diluting elevated control samples with a suitable reagent blank and must verify that exogenous analyte added to the controls can be quantitatively recovered. Suitable quality control samples should then be incorporated into the routine assays. After a minimum of 20 separate determinations, the mean and standard deviation (arithmetic or geometric as determined by distribution analysis) should be calculated for setting expected limits. This process is discussed in detail in a later section.

Thus, results on quality control samples are useful for documenting the limits of between-assay precision which can be obtained and for identifying analytical errors. The between-assay precision observed for below-normal, normal and above-normal control samples indicates relative degree of confidence which can be placed on an unknown (patient sample) result. In terms of the coefficient of variation, confidence in radioimmunoassay results is generally highest for sample results within the normal range and increasingly lower as results are further displaced from this region. Unacceptably low or high results obtained on all of the control samples imply an error in preparing standards.

Implementing the quality control recommendations described to this point is necessary to assure that tests are optimally performed in a given laboratory. A complete quality control program should also assess how the individual laboratory performs relative to other laboratories. Interlaboratory proficiency testing can provide such information. Since consensus standard materials for the various analytes measured by radioimmunoassay are scarce, data from interlaboratory surveys can provide a marker for a single laboratory's accuracy relative to that of other laboratories.

These quality control recommendations are not all inclusive but are intended to provide each laboratorian with a basis for designing a quality control program for any radioimmunoassay laboratory. References for the more detailed statistical and practical aspects of radioimmunoassay quality control are provided (*47,85,86*).

G. ENZYME-IMMUNOASSAY

1. Principles Involved

Enzyme-immunoassay (EIA) involves using immunologic techniques with enzyme-labeled reagents. The procedures used are similar to those used in radioimmunoassay. Antigen or antibody can be labeled and detected. Several assays have been developed, but few are widely used.

The two major divisions of EIA are: (1) homogeneous, in which bound and free fractions of the labeled reagent do not have to be separated, and (2) heterogeneous, in which bound and free labeled products must be separated to detect the immunologic reaction. There are several types of heterogeneous assays (*94*).

In homogeneous EIA, the enzyme label which is attached to an antigen is active, but when antibody binds to the antigen the enzyme becomes inactive, possibly as a result of steric interference. When unlabeled antigen (in standards or unknowns) is added to a labeled antigen-antibody complex, the unlabeled antigen competitively replaces the labeled antigen in the complex. Substrate is then added, and the enzyme (still attached to the antigen) can act on it to produce a quantifiable change. Homogeneous assays are readily automated and are used primarily in detecting small antigens such as drugs.

Heterogeneous EIA requires separating bound and free labeled reagent because the enzyme reacts the same in both states. Solid-phase assays are frequently used for simplicity. Four types of heterogeneous EIA have been developed (*122*).

(1) Competitive assay for antigen

A known amount of labeled antigen competes with the unlabeled antigen (standard or unknown) for sites on a known amount of antibody.

(2) Immunoenzymometric assay for antigen

The unlabeled antigen is added to a known amount of labeled antibody; the combined antibody is allowed to react with a solid phase unlabeled antigen.

(3) Sandwich

Unlabeled antigen is reacted with solid phase antibody; labeled antibody is added and reacts with the bound antigen.

(4) Enzyme linked immunoassay for antibody

Unlabeled antigen is attached to a solid phase; unknown antibody is allowed to react, and labeled antiglobulin is allowed to react with the bound antibody.

2. Problems Frequently Encountered

Some areas still to be worked out for EIA are (*34,94*):
(1) optimization of conjugate preparation,
(2) characterization and standardization of conjugates.
(3) quality control criteria for reagents,
(4) critical comparisons of various test principles with respect to sensitivity, precision, and practicality,
(5) potentials of alternatives to enzyme determinations such as fluorometry,
(6) automation,
(7) relationship between enzyme activity and antibody concentration, since it may not be linear.

When tubes are coated with antigen, the attachment is usually physical rather than covalent, and the antigen may leak during the testing procedure. This leakage may impair the sensitivity and precision of the test; therefore, the optimal conditions and amount of antigen coating must be determined for each assay. Temperature and concentration of nonspecific immunoglobulins in the serum may also affect sensitivity and precision of the assay. Cross-reactions and nonspecific reactions should also be controlled.

3. Control Recommendations

The required quality control measures for EIA have not yet been developed. Reference antigens and antisera should be used when applicable. Some criteria for selection of the enzyme label have been proposed (*34,94,122*). The enzyme should be:

(1) available cheaply in high purity,
(2) of high specific activity,
(3) stable under assay and storage conditions
(4) assayed by a method which is simple, sensitive, rapid, and inexpensive,
(5) soluble,
(6) absent from biological tissues, and
(7) capable of retaining activity during linkage reactions.
(8) Inhibitors and substrate must be absent from biological fluids.
(9) For homogeneous assays, the enzyme must be capable of inhibition or reactivation when antibody combines with the enzyme-antigen conjugate, and the enzyme assay conditions must be compatible with antigen-antibody binding.

H. ELECTROPHORESIS

1. Principles Involved

Electrophoresis is the movement of charged particles or molecules in solution under the influence of an electrical field. Because the rate of migration varies with different molecules, electrophoresis can be used to separate mixtures of compounds. The direction and speed of migration are determined by a number of interrelated factors. The nature of the particle (that is, its net charge, size, and frictional coefficient) is important. The buffer solution influences migration according to its pH, ionic strength, and viscosity. The intensity and direction of the electric current are also involved. The support medium (paper, gel, cellulose acetate) influences the degree of separation as well as the time required to separate particles adequately. All these factors may be altered to effect the desired separation (*39,64*).

Serum electrophoresis is usually carried out on cellulose acetate strips with a barbital buffer, pH 8.6. Five protein fractions are normally identified: albumin, and alpha-1, alpha-2, beta, and gamma globulins. Abnormal proteins may be seen if they are present in a sufficient concentration, and a decrease in concentration or the absence of a component may also be detected.

Variations in the absolute or relative amounts of the fractions are associated with a number of diseases.

The usefulness of electrophoresis has been expanded by combining it with immunoprecipitation. Some of the modifications of the combination are immunoelectrophoresis (IEP), immunoelectrodiffusion (IED), and counterelectrophoresis (CEP). These procedures are usually carried out in agar gel or cellulose acetate. Abnormalities in serum proteins can usually be detected with IEP. For this procedure the unknown serum is electrophoresed, and at the end of the run, antibody is applied along the length of the support medium so that the antibody and the separated fractions can diffuse toward each other. Precipitin arcs form where the protein and its specific antibody meet in optimal proportions. Monovalent or polyvalent antisera can be used, depending on the purpose of the test. The results obtained are interpreted on the basis of comparing the results of the unknown with those of a known control serum tested under the same conditions (*79*).

Another method of detecting serum components with electrophoresis and immunoprecipitation is counterelectrophoresis. This procedure makes use of endosmosis, which causes weakly charged negative molecules (the molecules are actually positively charged relative to the buffer) to move toward the cathode. When an antigen is strongly negative and moves toward the anode, and its corresponding antibody is weakly charged and moves toward the cathode, they may be placed in opposing wells so that they move toward each other and precipitate in the area between the wells. Known antigen can be used to detect antibodies in serum, (e.g., the test for histoplasmosis antibody) or known antibody can be used to detect antigens (e.g., the test for hepatitis B surface antigen).

A quantitative procedure involving electrophoresis and immunoprecipitation is IED or rocket electrophoresis. After the serum to be tested is placed in wells cut in a gel containing antibody, current is applied. As the serum migrates, the antigen to be quantitated reacts with the antibody in the gel to form a rocket-shaped area of precipitate. The length of the rocket is proportional to the concentration of the antigen in the serum. Specific values are assigned by comparing the length of the rocket with those of rockets formed with sera of known concentration (*28,64*).

2. Problems Frequently Encountered

The results of electrophoresis depend on the specimens used;

for example, a serum will give a different pattern from that obtained with plasma. Storage conditions may also affect the results. The expected results will depend on the patient's condition including age, medication and treatment, diet, and hereditary factors.

Deviations in patterns may result from inconsistent sample application, changes in buffer composition or pH, power supply fluctuations, alterations in staining procedures, and, if quantitated, changes in densitometer function (*40*). In addition, IEP results are affected by changes in antiserum (CEP also) and diffusion conditions. IEP, CEP, and IED may give inaccurate results if the gel has been allowed to dry at all before use.

3. Control Recommendations

Controls with known reactions or patterns should be included in each run, because results of these tests are frequently interpreted on the basis of a comparison with a normal. The control sera should be run enough times before comparison with unknowns so the person interpreting the reactions can recognize (qualitatively or quantitatively) when the procedure has varied.

The time of each run should be the same so that fractions or arcs can be identified reliably. Staining times should also be controlled. Reagents should not be used excessively because they become less efficient with repeated use. The pH of the electrophoresis buffer should be checked before use.

I. OTHER TESTS

1. Ox Cell Hemolysin (OCH) Test for Infectious Mononucleosis.

a. Principles Involved

During 1972-1978, CDC conducted a study comparing the OCH test, the heterophile test, and various IM slide tests (*21*). Although all three types of tests were satisfactorily reliable as diagnostic procedures for IM, CDC recommended that laboratories performing many tests use the OCH or heterophile test and that laboratories testing only a few specimens use an IM slide kit.

A greater specificity for IM is claimed with the OCH test (*59*) because of the occurrence of high-titered ox cell (beef) hemolysins in the sera of patients with IM. Normal sera contain very few ox cell hemolysins (*5*). Consequently, false reactions have been obtained with the OCH test only rarely, whereas false-negative and false-positive results may occur more frequently with the presumptive and differential heterophile protocols. The OCH test takes significantly less time to perform than do the presumptive and differential adsorption techniques.

b. Problems Frequently Encountered

In reading the OCH test, it may be somewhat difficult to match color intensity between test wells of a Microtiter plate and the 50% hemolytic standards. This problem occurs because a button of red cells remains in the bottom of a well (following centrifugation) and may interfere in matching attempts. If the test were performed in tubes, it would be easier to read because the standard could be juxtaposed with each dilution tube of the test specimen for a more direct comparison. However, an experienced laboratorian should be able to obtain accurate results using the microsystem. Although an excess of complement is added, severely anticomplementary sera may occasionally cause problems.

c. Control Recommendations

Controls should include known positive and negative sera as well as a cell control and a complement-cell control for the OCH test. The latter two controls should show no hemolysis.

d. Standardization

In the past, several technical difficulties were reported because of variations of test protocol. These included reporting titers based on final rather than serum dilutions, varying the amount of complement used, varying the incubation time, and diluting sera with saline containing $MgSO_4$ without buffer to control the pH. The standard protocol for the OCH test presently used at CDC (*76,115*) has done much to alleviate these problems.

IV. STANDARDIZATION

Lack of standardization is undoubtedly the greatest single factor contributing to the poor comparability of immunologic tests results today (*51*). Proficiency testing results have frequently been reported in which there was a greater than 1000-fold difference between the highest and lowest titers reported on a single serum sample (*109*). With the proliferation of methodologies and commercial test kits, standardization becomes imperative if laboratorians and physicians are to be able to interpret test results accurately.

Because it is difficult if not impossible to produce pure, homogeneous biological standards of known content similar to those that can be produced for chemical analyses, standardization in immunology usually takes one of two routes. One is to standardize the methodology, and the other is to prepare a physical reference material and assign arbitrary units to it.

It is often not practical or desirable to use a standard methodology in all laboratories for all applications of the test results. The reference methodology should serve only as the comparative standard to which other methods can be compared and by means of which results from different tests can be compared.

Although many antigens and antisera used in immunology are stable, they are frequently so complex and heterogeneous that it is difficult to produce high quality reference materials. Antigens are frequently mixtures of various bacterial or tissue components, and in addition, the antibodies to these mixtures are heterogeneous with respect to antibody class and subclass, antigen specificity, avidity, and affinity.

Standardization has been sponsored by international groups such as the World Health Organization (WHO), the International Union of Immunological Societies (IUIS), and by national organizations such as the National Committee for Clinical Laboratory Standards (NCCLS) and the Center for Disease Control (CDC). Table 3 lists standards that are available from various organizations.

Another approach to standardization is that of establishing product standards and evaluating products to see if they meet these standards. CDC has been involved in developing and applying such standards (*14, 15,104*). Table 4 shows the products for which such standards have been established.

Table 3. List of Available Standards

Analyte or Procedure*	Sponsoring Organization	Type**	Reference
Adenovirus	WHO	P	125
Alpha-fetoprotein	WHO	P	97,98
Amebiasis	CDC	P	16
Anthrax	WHO	P	125
Antinuclear factor	IUS/WHO/MRC	P,M	48,8,26,1
Antistreptolysin O	CDC/WHO	P,M	52,76,16,20,102
Arboviruses	CDC	P,M	16,76
Aspergillus	CDC	P	16
Autoimmunity	CDC	M	23
Bacterial agglutination tests	CDC	P,M	16,76
Bacterial identification	CDC	P	16
Blastomyces dermatitidis	CDC	P	16
Blood typing serum	WHO	P	125
Brucella	NADL/CDC/WHO	P,M	65,125,16,68,124,126
Calibration reference and contact materials (ASC-2)	NCCLS	M	70
Calibrations (PSC-5)	NCCLS	M	70
Candida albicans	CDC	P	16
Carcinoembryonic antigen	WHO	P	53
Chlamydia	CDC	P	16
Cholera	WHO	P	125
Coccidioides immitis	CDC	P	16
Colorado tick fever	CDC	P	16
Complement fixation	CDC/WHO	P,M	65,117
Coomb's test	FDA		38
Cortisol (hydrocortisone)	NBS	P	69
Coronavirus	CDC	P	16
Coxsackievirus	WHO/CDC	P	16,125
Cryptococcus neoformans	CDC	P,M	16,76
Cytomegalic virus	CDC	M	77
Dengue	CDC	P	16
Diphtheria	WHO	P	125
Echinococcus	WHO	P	125
Echovirus	WHO/CDC	P	125,16
Enterovirus	WHO	P	125
Equine encephalitis viruses	CDC	P	16
Evacuated tubes for blood specimen collection (ASH-1)	NCCLS	M	70
FITC	BSC	P	62
FITC-labeled antisera	MRC/WHO/CDC	P,M	48,16,26,1,63
Fluorescein/protein ratios in fluorescein isothiocyanate-labeled protein (TSM-4)	NCCLS	M	70
FTA-ABS	IUIS/WHO/CDC	P,M	48,8,16,26
Herpesvirus hominis HSV	CDC	P,M	16,77
Histoplasma capsulatum	CDC	P,M	16,76
Human chorionic gonadotrophin	WHO	P	125
Human chorionic gonadotrophin beta subunit	WHO	P	128
Human growth hormone	WHO	P	128
Human insulin	WHO	P	125,128
Human pituitary LH (ICSH)	WHO	P	128
Human TSH	WHO	P	128
Hyaluronidase	WHO	P	125
Immunoglobulin D	WHO	P	1,88
Immunoglobulin E	WHO	P	1,91,127
Immunoglobulin G, A, and M	WHO	P	79,49,90,89,13

Table 3. List of Available Standards – Continued

Analyte or Procedure*	Sponsoring Organization	Type**	Reference
Infectious mononucleosis	CDC	M	76
Influenza	WHO/CDC	P	125,16
Isotonic NaCl solution (TS1/BB-1)	NCCLS	M	70
Labeling efficiency of FITC (ASM-1)	NCCLS	M	70
Labeling of reagents (ASL-1)	NCCLS	M	70
Leptospira	WHO/CDC	P	16,125
Manuals for installation operation and repair of laboratory instruments (ASL-1)	NCCLS	M	70
Measles	WHO/CDC	P	16,125
Mumps	CDC	P	16
Mycoplasma pneumoniae	CDC/WHO	P,M	76,125,16
Opacity reference	WHO	P	81,125
Parainfluenza	WHO/CDC	P	16,125
Pertussis	WHO	P	125
Pneumococcus	WHO	P	125
Poliovirus	WHO/CDC	P	16,125
Power requirements (PSI-5)	NCCLS	M	70
Protein solution (ASC-1)	NCCLS	M	70
Pyrogen	WHO	P	125
Q fever	CDC	P	16
Rabies	WHO/CDC	P	16,125
Radioimmune systems (PSLA-1)	NCCLS	M	70
Reovirus	WHO/CDC	P	16,125
Respiratory syncytial virus	CDC	P	16
Rheumatoid factor	CDC/WHO	P,M	4,76,125
Rubella (TSM-5)	NCCLS/CDC/WHO	P,M	76,125,72,78,16,70
Service of laboratory instruments (PSI-6)	NCCLS	M	70
Skin puncture for specimen collection (PHS-4)	NCCLS	M	70
Smallpox	WHO	P	125
Spectrophotometric performance criteria (PSI-3)	NCCLS	M	70
Spotted fever group rickettsia	CDC	P	16
Staphylococcus	WHO	P	125
Storage and transport of specimens (PSH-5)	NCCLS	M	70
Syphilis	CDC/WHO	P,M	16,125
Temperature calibration (ASI-2)	NCCLS	M	70
Tetanus	WHO	P	125
Toxin identification	CDC	P	16
Toxoplasma gondii	CDC/IUIS/WHO	P,M	16,125,48,76,26
Trichinosis	CDC	M	76
Tuberculosis	WHO	P	125
Tularemia	CDC	M,P	16,76
Typhoid	WHO	P	125
Typhus group rickettsia	CDC	P	16
Vaccinia	CDC	P	16
Varicella/zoster	CDC	P	16
Venereal disease	CDC/IUIS/WHO	M,P	8,16,26,48,71
Viral identification	CDC	P	16
Water (TSC-3)	NCCLS	M	70
Venipuncture (PSH-3)	NCCLS	M	70
Yellow fever	CDC/WHO	P	16,125

* Numbers in parentheses refer to NCCLS publications

** P = Physical standard (Reference material)
M = Methodological standard (Reference method)

CDC = Center for Disease Control, Atlanta, GA
MRC = Medical Research Council, Division of Biological Standards, National Institute for Medical Research, Mill Hill, London, England
NCCLS = National Committee for Clinical Laboratory Standards, Villanova, PA
WHO = World Health Organization, Geneva, Switzerland
IUIS = International Union of Immunological Societies
FDA = Food and Drug Administration, Rockville, MD
NADL = National Animal Disease Laboratory, Ames, IA
BSC = Biological Stain Commission, Rochester, NY

Table 4. Products For Which Performance Specifications Have Been Established

GENERAL SPECIFICATIONS

Bacterial, Fungal, and Parasitic Antigens, and Animal Antisera (General)
Human Control Sera and Other Body Fluids (General)
Immunofluorescence Reagents (General)
Products for Nontreponemal Tests for Syphilis (General)
Multiple Unit Products (General)

BACTERIAL SPECIFICATIONS

GENERAL

Actinomyces israelii Fluorescein Isothiocyanate (FITC) Labeled Globulin
Actinomyces naeslundii Fluorescein Isothiocyanate (FITC) Labeled Globulin
Bordetella pertussis Fluorescein Isothiocyanate (FITC) Labeled Globulin
Bordetella parapertussis Fluorescein Isothiocyanate (FITC) Labeled Globulin
Brucella Agglutination Antigen
Brucella Antiserum
Brucellosis Control Human Serums
Francisella tularensis Antigen
Francisella tularensis Antisera
Tularemia Control Human Serums
Francisella tularensis Fluorescein Isothiocyanate (FITC) Labeled Globulin
Hemophilus influenzae Antiserum, Polyvalent
Hemophilus influenzae Antiserum, Monovalent
Hemophilus influenzae Fluorescein Isothiocyanate (FITC) Labeled Globulin, Polyvalent
Leptospiral Antisera
Neisseria meningitidis Antisera, Polyvalent
Neisseria meningitidis Antisera, Group-Specific
Neisseria meningitidis Fluorescein Isothiocyanate (FITC) Labeled Globulin, Polyvalent
Pseudomonas pseudomallei Fluorescein Isothiocyanate (FITC) Labeled Globulin
Staphylococcus Propagating Strains
Staphylococcus Typing Bacteriophages
Streptococcus pneumoniae Capsular Antisera, Pools
Streptococcus pneumoniae Capsular Antisera, Type-Specific
Streptococcus pneumoniae Fluorescein Isothiocyanate (FITC) Labeled Globulin, Polyvalent
Streptococcus Extracts, Group-Specific
Streptococcus Antisera, Group-Specific
Streptococcus Group A Fluorescein Isothiocyanate (FITC) Labeled Globulin
Streptococcus Group A Fluorescent Antibody Multiple Unit Products
Vibrio cholerae Antiserum, Polyvalent
Vibrio cholerae Antisera, Sub-O-Group-Specific

Table 4. Products For Which Performance Specifications Have Been Established (continued)

BACTERIAL SPECIFICATIONS

ENTERIC

Escherichia coli O Antisera, Polyvalent
Escherichia coli O Antisera
Escherichia coli O Antisera, Group or Sub-O-Group Specific
Escherichia coli K (KO) Antisera, Polyvalent
Escherichia coli K (KO) Antisera
Escherichia coli K (KO) Antisera, Adsorbed
Enteropathogenic *Escherichia coli* (EEC) Fluorescein Isothiocyanate (FITC) Labeled Globulin
Serogrouping and Pools
Escherichia coli Flagellar (H) Agglutination Antigen
Escherichia coli Flagellar (H) Antisera, Polyvalent
Escherichia coli Flagellar (H) Antisera
Escherichia coli Flagellar (H) Antisera, Adsorbed
Klebsiella Capsular Antisera, Pools
Klebsiella Slide Agglutination Capsular Antisera
Klebsiella Capsular Antisera, Type-Specific
Salmonella Somatic Antisera, Polyvalent
Salmonella Somatic Antisera, Group-Specific
Salmonella Somatic Antisera, Single Factor
Salmonella Vi Antisera
Salmonella Vi Fluorescein Isothiocyanate (FITC) Labeled Globulin
Salmonella Flagellar Antisera, Polyvalent
Salmonella Flagellar Antisera
Salmonella Flagellar Antisera, Single Factor
Salmonella OH Fluorescein Isothiocyanate (FITC) Labeled Globulin, Polyvalent
Salmonella OH Fluorescent Antibody Multiple Unit Products
Shigella Antisera, Type-Specific
Shigella Antisera, Subgroup-Specific
Shigella flexneri Fluorescein Isothiocyanate (FITC) Labeled Globulin, Polyvalent
Shigella sonnei Fluorescein Isothiocyanate (FITC) Labeled Globulin

VENEREAL DISEASE

VDRL Antigen
VDRL Buffered Saline
RPR Card Antigen Suspension
USR Antigen Suspension
Control Serums for Nontreponemal Tests for Syphilis
Treponema pallidum Antigen for FTA-ABS Test
Anti-human Globulin, Fluorescein Isothiocyanate (FITC) Labeled, for the FTA-ABS Test for Syphilis
FTA-ABS Test Sorbent
Control Serums for FTA-ABS Test
FTA-ABS Test Multiple Unit Product
Treponema pallidum Fluorescein Isothiocyanate (FITC) Labeled Globulin
Neisseria gonorrhoeae Fluorescein Isothiocyanate (FITC) Labeled Globulin

Table 4. Products For Which Performance Specifications Have Been Established (continued)

FUNGAL AND PARASITIC
Aspergillus Immunodiffusion (ID) Antigens
Aspergillus Immunodiffusion (ID) Antisera
Aspergillosis Positive Control Human Serums
Blastomyces dermatitidis Yeast Form Complement Fixation (CF) Antigen
Blastomyces dermatitidis Yeast Form Immunodiffusion (ID) Antigen
Blastomycosis Positive Control Human Serum
Blastomyces dermatitidis Yeast Form Fluorescein Isothiocyanate (FITC) Labeled Globulin
Coccidioides immitis (Coccidioidin) Complement Fixation (CF) and/or Immunodiffusion (ID) Antigen
Coccidioidomycosis Positive Control Human Serum
Coccidioides immitis Tissue Form Fluorescein Isothiocyanate (FITC) Labeled Globulin
Cryptococcus neoformans Latex Agglutination (LA) Reagent
Cryptococcus neoformans Latex Agglutination (LA) Control Reagent
Cryptococcus neoformans Polysaccharide Antigen-Positive Human Body Fluid
Cryptococcus neoformans Fluorescein Isothiocyanate (FITC) Labeled Globulin
Histoplasma capsulatum Mycelial Form Complement Fixation (CF) and/or Immunodiffusion (ID) Antigen
Histoplasma capsulatum Yeast Form Complement Fixation (CF) Antigen
Histoplasmosis Positive Control Human Serum
Histoplasma Capsulatum Yeast Form Fluorescein Isothiocyanate (FITC) Labeled Globulin, Polyvalent
Sporothrix schenckii Yeast Form Fluorescein Isothiocyanate (FITC) Labeled Globulin
Toxoplasma gondii Indirect Fluorescent Antibody (IFA) Antigen
Toxoplasmosis Control Human Serums
Anti-human Fluorescein Isothiocyanate (FITC) Labeled Globulin, Toxoplasmosis Indirect Fluorescent Antibody (IFA) Test
Toxoplasmosis Indirect Fluorescent Antibody (IFA) Diagnostic Multiple Unit Products

Table 4. Products For Which Performance Specifications Have Been Established (continued)

CHLAMYDIAL/MYCOPLASMAL/RICKETTSIAL/VIRAL ANIMAL ANTISERA AND IMMUNE ASCITIC FLUIDS
Animal Antisera and Immune Ascitic Fluids (General)
Negative or Pre-Immune Animal Serum
Adenovirus Typing
Adenovirus Group
California (LaCrosse)
Chlamydia Group
Colorado Tick Fever
Coronavirus
Coxsackievirus Typing, Type A
Coxsackievirus Typing, Type B
Cytomegalovirus
Eastern Equine Encephalomyelitis
Echovirus Typing
Epidemic Typhus Fever
Herpesvirus Hominis
Herpesvirus hominis Type 1 or 2 for Neutralization
Influenza Strain Specific
Influenza Typing
Measles (Rubeola)
Mumps
Murine Typhus Fever
Mycoplasma pneumoniae
Newcastle Disease
Parainfluenza
Poliovirus Typing
Q Fever
Reovirus Typing
Respiratory Syncytial
St. Louis Encephalitis
Spotted Fever Group
Vaccinia
Varicella-Zoster
Venezuelan Equine Encephalomyelitis
Western Equine Encephalomyelitis
Yellow Fever

BEST DOCUMENT AVAILABLE

Table 4. Products For Which Performance Specifications Have Been Established (continued)

HUMAN IMMUNE SERA AND PLASMAS
Human Immune Sera and Plasmas (General)
Nonreactive Human Sera and Plasmas
Adenovirus
California Encephalitis
Chlamydia Group
Colorado Tick Fever
Coronavirus OC43
Coxsackievirus, Type A or B
Cytomegalovirus
Eastern Equine Encephalomyelitis
Epidemic Typhus Fever
Herpesvirus hominis
Influenza Type A or B
Measles (Rubeola)
Mumps
Murine Typhus Fever
Mycoplasma pneumoniae
Parainfluenza
Poliovirus
Q fever
Respiratory Syncytial
Rocky Mountain Spotted Fever
St. Louis Encephalitis
Vaccinia
Varicella-Zoster
Venezuelan Equine Encephalomyelitis
Western Equine Encephalomyelitis
Yellow Fever

Table 4. Products For Which Performance Specifications Have Been Established (continued)

ANTIGENS
Antigens (General)
Normal Tissue Control
Adenovirus, CF and HAI
California (LaCrosse), CF and/or HAI
Chlamydia Group, CF
Colorado Tick Fever, CF
Coronavirus OC43, CF and/or HAI
Coronavirus 229E, CF
Cytomegalovirus, CF and/or IHA
Eastern Equine Encephalomyelitis, CF and/or HAI
Echovirus, HAI
Herpesvirus hominis, CF and/or IHA
Influenza Ribonucleoprotein, CF
Influenza, HAI
Measles (Rubeola), CF and/or HAI
Mumps, Viral, CF and/or HAI
Mycoplasma pneumoniae, CF
Mycoplasma pneumoniae, IHA
Newcastle Disease, HAI
Parainfluenza, CF and/or HAI
Poliovirus, CF
Q Fever, CF
Reovirus, HAI
Respiratory Syncytial, CF
Rubella, HAI
St. Louis Encephalitis, CF and/or HAI
Simian Virus 40, HAI
Spotted Fever Group, CF
Typhus Fever Group, CF
Vaccinia, CF and/or HAI
Varicella-Zoster, CF
Venezuelan Equine Encephalomyelitis, CF and/or HAI
Western Equine Encephalomyelitis, CF and/or HAI
Yellow Fever, CF and/or HAI

Table 4. Products For Which Performance Specifications Have Been Established (continued)

MISCELLANEOUS REAGENTS AND MULTIPLE COMPONENT SYSTEMS
FLUORESCENT ANTIBODY REAGENTS
Immunofluorescent Reagents (General)
Rabies Mouse Brain Impression Slides
Normal Mouse Brain Impression Slides
Rabies Mouse Brain Adsorbing Suspension
Normal Mouse Brain Adsorbing Suspension
Fluorescein Isothiocyanate (FITC) Labeled Anti-Rabies Globulin
MULTIPLE UNIT PRODUCTS
Multiple Unit Product (General)
Rubella Control Sera, HAI
Rubella Multiple Unit Product, HAI
OTHER REAGENTS
Guinea Pig Complement
Hemolysin (Sheep Cell Antiserum)
Receptor Destroying Enzyme

METHODS OF EVALUATION
Tube Agglutination Test
Slide Agglutination Test
Quellung Test
Fungal Micro-Immunodiffusion Test
Standardization of Erythrocyte Suspensions by Spectrophotometer
Laboratory Branch Complement Fixation Test (LBCF)
Cryptococcus Latex Agglutination Test
Standardized Hemagglutination (HA) and Hemagglutination-Inhibition (HAI) Test
Use and Evaluation of Receptor Destroying Enzyme (RDE)
Arbovirus Hemagglutination (HA) and Hemagglutination-Inhibition (HAI) Test
Rubella Hemagglutination (HA) and Hemagglutination-Inhibition (HAI) Test
Indirect Hemagglutination (IHA) Test
Neutralization Index (NI)
Neutralization (Nt) Tests
Rabies Fluorescent Antibody Test
Metabolic Inhibition Test for Mycoplasma

Table 4. Products For Which Performance Specifications Have Been Established (continued)

PREPARATION OF EVALUATION MATERIALS

- McFarland Density Standards
- Leptospiral Microscopic Agglutination Test
  - Phosphate Buffered 0.85% NaC1, pH 7.5-7.6
- *Shigella* Slide Agglutination Test
  - 0.85% NaC1 Solution Containing Mercuric Iodide
- Fungal Micro-Immunodiffusion (ID) Test
  - Phenolized Agar Medium for *Aspergillus* ID Test
  - Phenolized Agar Medium for *Blastomyces, Coccidioides,* and *Histoplasma* ID Tests
- Complement Fixation Test
  - Barbital Buffered Diluent (BBD)
- *Cryptococcus* Latex Agglutination Test
  - Glycine Buffered 0.9% NaC1 Solution, pH 9.0
- Immunofluorescence Tests
  - 0.01M Phosphate Buffered Saline (PBS), pH 7.2-7.3
  - 0.01M Phosphate Buffered Saline (PBS), pH 7.6-7.8
  - Carbonate-Bicarbonate Buffer, pH 9.0
  - Buffered Glycerol Mounting Medium, pH 9.0
  - Trypticase Soy-Tryptose Broth for Salmonellae
  - 0.85% NaCl Solution Containing 0.6% Formalin
  - Kirkpatrick's Fixative for *Salmonella*
  - Anaerobic Culture Medium for Actinomycetes
  - 0.2% Evans Blue Solution for Toxoplasmosis IFA Test
- Erythrocyte Preservatives
  - Alsever's Solution
  - Sodium Citrate (3.8%)
  - Dextrose Gelatin Veronal
- Standardized Hemagglutination and Hemagglutination-Inhibition Test
  - Calcium Saline Solution
  - Sodium Citrate (2.5%)
  - Phosphate Buffered Saline
- Arbovirus Hemagglutination and Hemagglutination-Inhibition Test
  - Borate Saline
  - Kaolin Solution (25%)
  - Bovalbum (4%)
  - Phosphate Buffers for Goose Erthrocytes (0.2M)
- Rubella Hemagglutination and Hemagglutination-Inhibition Test
  - Hepes-Saline-Albumin-Gelatin-Diluent
  - Heparin-MnCl$_2$ Reagent
- Indirect Hemagglutination Test
  - Phosphate Buffered Saline
  - Normal Horse Serum Diluent
  - Normal Rabbit Serum Diluent for CMV and Herpes
- Neutralization Tests
  - Hank's Balanced Salt Solution
- Rabies Fluorescent Antibody Test
  - Phosphate Buffered Saline, pH 7.6-7.8
  - Carbonate Bicarbonate Buffer, pH 9.0
  - Buffered Glycerol Mounting Medium, pH 9.0

V. EXTERNAL QUALITY CONTROL

External quality control consists of applying control procedures that originate outside the laboratory being tested. External quality control includes exchanging samples with other laboratories, testing "check samples," and participating in formal proficiency testing programs. External controls permit evaluation of interlaboratory comparability and are conducive to standardization. Comparison of the results from different laboratories sometimes allows errors to be detected that may be overlooked with internal quality control programs. In addition, proficiency testing results can be used to evaluate the performance of methods and detect inadequacies that might not otherwise be found.

Many laboratories participate in proficiency testing programs because of regulatory requirements (Clinical Laboratories Improvement Act of 1967 or state regulations), but others participate voluntarily. In the future legislation may be expanded to include most laboratories, even those operated in physicians' offices.

Various methods of testing a laboratory's proficiency by submitting "unknown" specimens have been used for a long time (*9,41*). In addition to the traditional method of using mailed-in samples, several other approaches have been used or considered (*18,41,44,121*). These include (1) "on-site" proficiency testing, in which an examiner hand carries the samples into the laboratory and observes the testing procedure; (2) "blind" proficiency testing, in which the cooperation of the user of the laboratory service is enlisted to submit a sample which is disguised as a patient sample; (3) "reverse" proficiency testing, in which previously tested patient samples are picked up from the laboratory and subsequently tested by reference laboratories; and (4) "patient population" proficiency testing, in which analyses of the routine patient test results are used to monitor proficiency.

Experience with proficiency testing indicates that the mailed-sample proficiency testing is most useful to laboratories interested in self-improvement. If the samples are tested by the people who would ordinarily test the patient samples by the procedures that are usually used, the results are more meaningful. Under these circumstances errors might be detected that might be missed with the internal quality control program. During on-site examinations, laboratorians and examiners can exchange input and suggestions. Because the examiners are familiar with the problems in other laboratories they are frequently valuable sources of information and assistance.

Proficiency testing programs provide data on how well individual laboratories perform, but they also can provide comparisons of results obtained with various methods. An estimate of the sensitivity and specificity of particular kits or products can be determined from proficiency testing data.

Comparisons can be made by method or without regard to method. These comparisons give the laboratorian an idea of how his results compare with those of all other participating laboratories and how they compare with those obtained by other laboratories using the same method.

When sensitivity and specificity are estimated in proficiency testing programs, the results are not as accurate or reliable as when they are obtained in evaluation surveys, but they do provide valuable information that is otherwise not available. The more desirable evaluation survey is rarely used because of attendant logistic problems and its extremely high cost (*105,106,116*).

Sensitivity and specificity in the context of proficiency testing data refer to the ability of a test to give a positive or negative result on a particular sample with an accurately measured level of analyte or with a defined-cross-reacting or interfering component. If this type of information is obtained for a number of analyte levels and possible false-reaction producing conditions and is correlated with the frequency that these levels and conditions exist in patient populations, an idea can be obtained of the effectiveness of these tests (what their sensitivity and specificity would be with that population).

Proficiency testing data also provide such useful information as distribution of test use and comparison of test performance in the laboratory that can be used in selecting tests or in correcting problems that occur. Evaluations of test performance derived from proficiency testing data differ from the more commonly reported evaluations in that the former more nearly reflect the kind of performance that actually occurs in day-to-day practice. Most evaluations are conducted in fairly specialized facilities such as large hospital, public health, or university laboratories, and are frequently performed by personnel with special training and interest in the tests being evaluated. On the other hand, proficiency testing results come from a larger variety of personnel who are more likely to be providing the same types of results for patients. Test results are known to be more precise in the "hands of the devoted" than when others use a particular method (*121*).

Proficiency testing samples are frequently paired, and repeatability and reproducibility can thus be calculated. Because the number of paired sera is low for most analytes in most programs, it may not be possible to make an accurate estimate of repeatability and reproducibility in a participant laboratory, but at least an indication of problems can be obtained. If repeatability and reproducibility estimates from proficiency testing surveys are considered in conjunction with results from internal quality control, a more reliable picture can be obtained.

Proficiency testing results provide comparisons of the reproducibility that is achieved by method and by procedural variation. This information is valuable in selecting tests and in evaluating laboratory performance. It gives personnel an idea of how their laboratory's performance compares with that of other laboratories using the same methods and an idea of how the various procedural variations may affect reproducibility. Performance by method gives an indication of the inherent reproducibility of particular test methods.

Participation in proficiency testing programs promotes standardization in several ways. Participants learn what tests are commonly used and are thereby encouraged to adopt better and more standard tests. Standardization of reporting units is also facilitated by proficiency testing because of the necessity of minimizing the number of different units reported. Probably the most important mechanism is that of providing materials that can be used to calibrate and standardize test results. These materials can be the proficiency testing samples themselves or can be actual reference material (*36,109*).

Proficiency testing samples provide excellent reference material (survey-validated samples) because they have been analyzed in many laboratories and by many methods.

Proficiency testing can also make valuable contributions to improving the quality of laboratory test results by its educational function (*22,29,80*). In addition to the educational benefits that have already been discussed, proficiency testing programs can provide other information in the form of summary analyses, critiques, laboratory manuals, and publications in professional journals and other literature. A source of information that is provided by most proficiency testing programs but which is often overlooked by participants is that of consultation. Most proficiency testing agencies have well-qualified, experienced personnel available to assist in determining the sources of problems and to make suggestions toward correcting them. These agencies and the commercial organizations which provide the equipment and reagents should be consulted when problems arise.

Proficiency testing programs have been used to determine which components of the overall quality control program are important and how much effect they have on the final outcome (*80,105,106*).

VI. QUALITY CONTROL CALCULATIONS

Laboratory measurements are used by physicians for two purposes: (1) to identify diseased individuals, and (2) to follow the progress of diseased individuals during therapy. The laboratory itself has two responsibilities to the physician: (1) to provide the physician with an estimate of the variation in the normal population, and (2) to guarantee the reliability of individual measurements (*27*).

In immunology two different approaches are necessary in identifying diseased individuals: (1) individuals with a certain condition must be separated from the normal population in which the condition does not usually exist, and (2) individuals with abnormal levels of a particular analyte must be separated from those with normal levels. The first approach usually involves qualitative tests, and the laboratories' main responsibilities are: (1) to provide the physician with results that are derived from sensitive and specific tests, and (2) to provide an estimate of how frequently and under what circumstances the condition exists in the normal population. The quality control measures are used to verify that the test, as used, can detect all the positives without giving any false-positive results. The second approach involves quantitative and semiquantitative tests and can rely on the same methods that are used for other clinical laboratory analyses, except that the frequency distributions in immunologic assays are often log-normal rather than normal.

The sensitivity of a test in detecting a disease condition can be determined by measuring the prevalence of positive reactions in populations known to be positive for the condition. The specificity of the test is determined by measuring the frequency of negative reactions in populations known to be free of the condition including populations in which false-positive results might be obtained. This information and reproducibility data provide laboratory evidence of the disease state of the patient.

Even though the Gaussian law of errors was originally applied to repeated measurements of the same object and not to a series of measurements on different objects (*32,37*), empirical evidence shows that the distribution of a series of measurements on different objects often conforms to the Gaussian or normal distribution or some simple transformation of that distribution. This patterning has served as the basis of the commonly used measures of normal limits.

The distribution of values for an analyte in the normal population can be determined and described as follows: (1) Define exactly what the population is, i.e., sex, age, geographic location, health status. (2) Select a representative (random) sample from the population, and measure the analyte levels in that sample. (3) Determine whether the distribution fits the Gaussian distribution or some transformation of it (Kolmogorov-Smirnov test (*60,129*), Section VI B.2). (4) If the distribution is Gaussian or can be transformed to Gaussian, describe the distribution by calculating the mean and standard deviation of the appropriately transformed data.

The values in a normal population can be completely and accurately characterized by the mean and standard deviation, thus allowing estimates of the probability that an observed value is from that population. This information and that on the accuracy and precision of the determination provide laboratory evidence of the disease state of the patient and/or the efficacy of therapy.

It is important to point out that evaluations of tests and populations are only valid for the conditions and populations tested and cannot be extrapolated to other conditions and populations. Previous evaluations and determinations can be used as a guide, but reliable interpretation of test results must be based on data obtained under the same testing conditions for the same population.

A. INTRODUCTION TO STATISTICAL CALCULATIONS WITH GEOMETRIC DATA

Since the results of immunologic tests frequently do not fit a normal distribution, the usual arithmetic statistical formulas to determine means and standard deviations cannot be usefully applied. However, if the results are first converted to logarithmic form, a normal distribution is obtained, and the standard formulas will provide the desired information (*103,107,108*). Data which fit a normal distribution only after logarithmic conversion are called geometric and occur when there is a common ratio rather than a common difference between pairs of successive possible values. For example, in a dilution series based on twofold dilutions starting with 1 to 20, the common ratio is twofold, that is, 20 to 40, 40 to 80, 80 to 160, and 160 to 320 (ratios = 1 to 2). This ratio can be an inherent property of the factor being measured (multiplication of cells, immunoglobulin levels), or it can be the result of the measuring procedure used, as in the use of serial dilutions. Methods for making the necessary statistical calculations to aid in test selection, to establish a quality control system, and to assist in interpreting immunologic (geometric) data are described in this section.

Usually little or no mention is made of the statistical treatment of geometric data either in statistical texts or in laboratory manuals and texts. As a result, laboratorians have generally failed to require that the same standards be applied when geometric data are generated as are applied in other areas of the laboratory or inappropriate statistical tests have been used. Frequently, the assumption is that serologic tests are "accurate to within one tube," and that a difference of "two tubes" is a significant difference. This assumption is vague and may not be valid, depending on many variables. But, if acceptable limits (95% confidence limits) are set at the reported titer "plus or minus one tube," with a twofold ratio between tubes, a fourfold rise in titer on paired acute-phase convalescent-phase sera would not be a statistically significant change, because such variation would be within the acceptable limits for each specimen if the true titers were identical. For example, if the true brucella titers of both the acute-phase and convalescent-phase sera from a patient were 160, a titer of 80 measured on the acute-phase specimen would be within the acceptable limits, and a titer of 320 on the convalescent-phase serum would also be within the acceptable limits. This fourfold titer ratio would commonly be interpreted as a significant rise when in fact it might represent no change.

The variability in immunologic tests can and should be monitored with the same methods applicable to other quantitative tests. Results of repeated testing of control sera should be accumulated until there are enough to allow reasonably reliable calculations of the geometric mean and standard deviation (generally at least 30 results). The results obtained for each control serum (high titer or concentration and low titer or concentration) should then be used to construct quality control charts similar to those recommended for other areas of the laboratory.

The charts should indicate the mean, the upper and lower acceptability limits, and should provide for the routine plotting of control results so that shifts, trends, and "out-of-control" results can be readily detected.

Control sera should be available in sufficient quantity to last 6 months or more (if possible) to provide continuous monitoring as reagents or kits change.

Before meaningful information can be obtained by using arithmetic formulas to calculate means and standard deviations for geometric data, the data must first be converted to logarithmic form. Logarithms to any base may be used, but when doubling dilutions are used, logarithms to the base 2 are convenient.

The necessary formulas for calculating the arithmetic mean and standard deviation are:

Mean

$$\bar{x} = \frac{\Sigma x}{n}$$

Standard Deviation

Calculator Formula*

$$S = \sqrt{\frac{\Sigma x^2 - \frac{(\Sigma x)^2}{n}}{n-1}}$$

Manual Formula*

$$S = \sqrt{\frac{\Sigma (x-\bar{x})^2}{n-1}}$$

Where:
$\bar{x}$ = mean
n = number of results
S = standard deviation
Σx = the sum of all x values
Σx^2 = the sum of all x values squared

To calculate the log of the geometric mean of the three results 2, 4, and 8, change the results to logarithmic form, which would be 1, 2, and 3, respectively (that is, 2^1, 2^2, and 2^3) and use the arithmetic formula which has been changed to indicate the log conversion

$$\log \bar{x}_G = \frac{\Sigma \log x}{n}$$

$$\log \bar{x}_G = \frac{1+2+3}{3}$$

$$\log \bar{x}_G = 2$$

The geometric mean ($\bar{x}_G$) is then the antilog of 2, which is 4 ($2^2 = 4$).

The log of the geometric standard deviation (S_G) can be calculated by the *manual method* as follows:

---

*The calculator formula for computing the standard deviation is easier to use than the manual formula if a calculator is available that can accumulate Σx and Σx^2 simultaneously. The manual formula may be easier to use if a calculator with memories is not available.

$\log x$	$\log \bar{x}_G$	$(\log x - \log \bar{x}_G)$	$(\log x - \log \bar{x}_G)^2$
1	2	-1	1
2	2	0	0
3	2	1	1
$\Sigma \log x = 6$		$\Sigma (\log x - \log \bar{x}_G) = 0$	$\Sigma (\log x - \log \bar{x}_G)^2 = 2$

$$\log S_G = \sqrt{\frac{\Sigma (\log x - \log \bar{x})^2}{n-1}}$$

$$\log S_G = \sqrt{\frac{(1-2)^2 + (2-2)^2 + (3-2)^2}{3-1}}$$

$$\log S_G = \sqrt{\frac{1+0+1}{2}}$$

$$\log S_G = \sqrt{1}$$

$$\log S_G = 1$$

$$S_G = 2$$

The range of one geometric standard deviation is expressed as the geometric mean multiplied by (times) or divided by (x/÷) one standard deviation rather than plus or minus (±) one standard deviation since the intervals in a geometric distribution are characterized by a common ratio. The one geometric standard deviation range includes all values between the geometric mean divided by the geometric standard deviation and the geometric mean multiplied by the geometric standard deviation. To calculate ranges other than one geometric standard deviation, the geometric standard deviation is raised to the power of the number of geometric standard deviations desired. For example, the two geometric standard deviation range is $\bar{x}_G$ x/÷ $(S_G)^2$ and the three geometric standard deviation range is $\bar{x}_G$ x/÷ $(S_G)^3$. For the above example the 2 S_G range is 4 x/÷4 and the 3 S_G range is 4 x/÷8.

Negative results and zero values present a special problem in the analysis of geometric data because logarithms are defined only for positive numbers. One solution to the problem of zero values is to add 1 to each result, perform the calculations, and then subtract 1 from the mean. For example, to calculate the geometric mean of the titers (negative or zero, 10, and 20), add one to each value so they become 1, 11 and 21. Calculate the sum of their logarithms_{10} (0 + 1.04 + 1.32 = 2.36); divide by n (2.36 ÷ 3 = 0.788); determine the antilog_{10} (antilog_{10} 0.788 = 6.14), and subtract one to get the mean (5.14). The S_G obtained does not require correction.

Under some circumstances it would be more appropriate to consider the negative results as having titers one dilution level below the lowest level determined in the titration system (i.e., titers of < 8 treated as 4 in statistical calculations). If titers of 8, 16, 8 and < 8 were obtained on the same serum, the latter method would be more appropriate.

Although the mean, upper limit, or lower limit may not coincide with an obtainable titer, this fact does not limit the usefulness or the validity of the quality control procedures. A quality control chart could be prepared as in Figure 1.

When titers are obtained by multiple testing, and the log values of the averages (means) do not correspond to an obtainable titer, it may be necessary or desirable to calculate titers which are "between tubes" or are fractional logs. The formula for changing from $\log_a$ to $\log_b$ is:

$$\log_b x = \frac{\log_a x}{\log_a b}$$

where x = the value of the parameter, a = the base of the original data, and b = the base to which the parameter is to be changed. For example to change from $\log_{10}$ to $\log_2$:

$$\log_2 x = \frac{\log_{10} x}{\log_{10} 2} = \frac{\log_{10} x}{.301}$$

and to change back to $\log_{10}$:

$$\log_{10} x = \frac{\log_2 x}{\log_2 10} = \frac{\log_2 x}{3.322}.$$

This can also be done by referring to log tables or by using a calculator which contains a log function, but log tables for the bases that are commonly used in serology are often not available (log base 2, for example). Conversion factors can be used to change logs from one base to another. The following table gives some commonly used conversion factors:

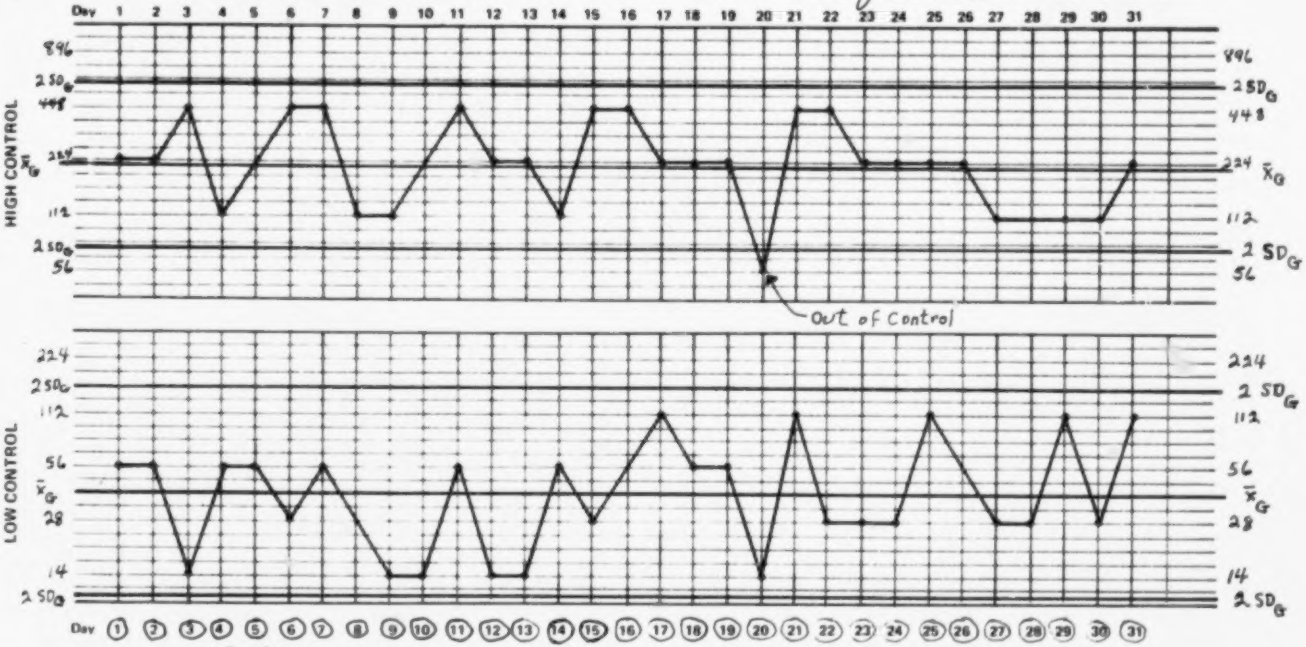
Table 5. Log Base Conversion Table

From A ↓ / To B →	1.41*	2	e	10
1.41*	–	0.500	0.347	0.151
2	2.000	–	0.693	0.301
e	2.885	1.443	–	0.434
10	6.644	3.322	2.303	–

*1.41 = $\sqrt{2}$, which is the base of the CDC ASO procedure. It consists of two parallel doubling dilution series, so that the first tube of the second series falls midway between the first and second tubes of the first series.

Figure 1. Immunology Quality Control Chart

ANALYTE Infectious Mononucleosis



Quality Control Calculations

MONTH July

Circle day when negative control is negative (if applicable).

	HIGH CONTORL	LOW CONTROL
$\bar{x}_G$	214	39
SD_G	1.76	1.99
LIMITS	69-661	10-154

To convert from logs to the base A (left hand column) to logs to the base B (top row), multiply the base A logarithm by the conversion factor at the intersection of the appropriate row and column. The same procedure is used to convert titers and standard deviations.

Because only certain values are obtainable in most serologic determinations, situations may arise in which no variation is acceptable (within 95% confidence limits). If close to 95% of the results are the same, any other titer will be beyond the two standard deviation limits and will be unacceptable.

For example, in the case illustrated in Figure 2A in which 83% of the results were 2000, and 17% were 4000, the acceptable 95% limits included all titers between 1328 and 3794. Titers of 1000 or 4000 are beyond those acceptable limits, consequently, only titers of 2000 are acceptable using the dilution scheme.

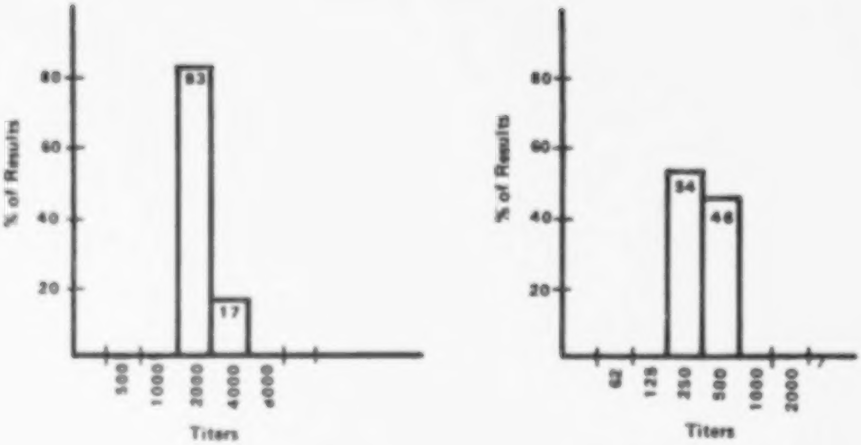
If fewer identical results are obtained with a particular dilution scheme, then dilutions above and below the median will be acceptable depending on the degree of variation exhibited. In addition, if the true titer is intermediate between the two possible titers and if 95% of the determinations consist of those two results, then they will be the only acceptable results.

As an example, suppose 54% of the results were 250, and 46% were 500. As shown in Figure 2B, the acceptable (95%) limits include all titers between 168 and 622. That means that only titers of 250 and 500 are acceptable if this dilution scheme is used. If other dilution schemes are used, any titer between 168 and 622 is acceptable.

When results from different laboratories or procedures are compared and different dilution schemes are used, it may be erroneous to assume that equivalent tube numbers represent equivalent titers. For example, if a twofold dilution scheme starting with 1:8 is used in one laboratory, and one of 1:10 is used in another laboratory, a titer of 16 obtained on the first laboratory is not equivalent to a titer of 20 obtained in the second laboratory (the second value is 25% higher than the first). Regardless of the dilution scheme used, the geometric means calculated from results of repeated testing of a single serum should be the same in both laboratories if the test procedures of the two laboratories are comparable and if both schemes permit possible results clustering on either side of the mean.

Instead of reporting results of immunologic tests as titers only, it would be better to report the titers with the error limits. Then the

Figure 2. Distributions of Test Results



A. $\bar{X}_G = 2245$, $SD_G = 1.30$

Acceptable limits- 1328 to 3794

B. $\bar{X}_G = 344$, $SD_G = 1.43$

Acceptable limits- 168 to 622

results could be interpreted and judgments made on the basis of the measured accuracy of the test procedure, rather than on the basis of undocumented general assumptions. When acute-convalescent titer changes are reported, it would be better to indicate the significance of the observed difference in titers rather than to report titers only or to rely on the traditional dictum that a fourfold increase is a significant change.

Immunologic data are amenable to the same statistical procedures as other quantitative and semiquantitative data and should be analyzed with those techniques. This treatment of the problems of statistical analysis of immunologic data should demonstrate that statistical analysis can be applied to such data. Application of these statistical measurements to immunologic data in the laboratory has revealed that the commonly held assumptions concerning the variability of immunologic test results are vague and inaccurate.

B. TESTS OF GOODNESS OF FIT

One of the first steps in the proper statistical analysis of data is to determine the probable distribution of the observed data so that the appropriate statistical tests can be applied. If the results are normally distributed, the common parametric methods described in statistical texts and in many laboratory references are appropriate. If the results have a significantly skewed distribution but can be transformed to a Gaussian distribution, the same methods can be used. If neither of these conditions exists, it is necessary to use nonparametric methods for analysis.

The probable distribution is determined by hypothesizing a specified distribution and using statistical tests to determine whether there is significant evidence against the hypothesis. This is usually stated in terms of a null hypothesis (H_0), i.e., that the assumed frequency distribution is the same as the sample frequency distribution and that any difference observed results from random variation. The Chi-square (χ^2) and Kolmogorov-Smirnov (K-S) tests are useful in making these determinations.

1. Chi-Square Test

This test is used to compare the actual numbers of observations and the expected (under the assumption) numbers of observations for the different values (classes) of the variant. In other words, the actual number of results in each range of the total distribution of results is compared to the number that would be expected if the hypothesized frequency distribution were the true frequency distribution. The Chi-square statistic is calculated as the sum of the squares of the differences between the observed and expected frequencies divided by the expected frequency.

$$\chi^2 = \Sigma \frac{(O - E)^2}{E}$$

where O = number of observations;

E = expected numbers which are calculated by using the assumed distribution.

The "degrees of freedom" for the chi-square test is the number of classes (k) minus the number of parameters estimated by sample statistics (p), minus one (d.f.= k-p-1). An example involving the normal distribution is studied next using proficiency testing data on IgG.

Since we are dealing with immunologic data, we assume a log-normal distribution, so a log transformation is necessary before we apply the test. The distribution of quantitative results on IgG is shown in Table 6. The expected frequencies (column 8 in Table 6) can be derived from normal distribution tables such as shown in Appendix A.

The χ^2 value is determined as follows (refer to Table 6):

Step 1. List the class limits as shown in column 1.

Step 2. List the class midpoints as shown in column 2.

Step 3. Convert the class limit values to logarithms to any convenient base (base e was used in this example).

Step 4. The geometric mean and standard deviation (arithmetic mean and standard deviation of the log transformed values) are calculated by the methods shown in part A of this section. The values in column 2 and the frequencies in column 7 produce sample estimates of $\overline{x}_G = 942$ and $S_G = 1.157$ or $\overline{x} = 6.8481$ and $S = 0.1454$ after log transformation.

Step 5. Calculate the class limits in terms of a z score. The z score is simply the distance from the mean to the x value expressed as number of standard deviations. It is calculated by the formula:

$$z = \frac{x - \overline{x}}{S}$$

The first value in column 4 was obtained as follows:

$$z = \frac{x - \overline{x}}{S} = \frac{6.3099 - 6.8481}{0.1454} = 3.70$$

Step 6. The area between zero and the z score is determined by referring to a table of normal frequency distributions (Appendix A). The values obtained are listed in column 5.

Step 7. Calculate the area within each class by subtracting the lower limit area listed in column 5 from the upper limit area listed in that column. For example the area in the class of 550 to 600 is:

$$(-0.4990)-(-0.4999) = 0.0009.$$

This is the probability that a result from this frequency distribution will occur in this class.

Step 8. Calculate the expected frequencies (column 8) by multiplying the probability that a result will occur in each class (column 6) by the total number of results. The first three rows in columns 7 and 8 are combined because the frequencies are too small. Ostle states that for small expected frequencies, the Chi-square statistic will be a poor indicator of the validity of the hypothesis under test (73).

Step 9. Calculate the Chi-square values for each class by the formula:

$$\chi^2 = \frac{(0 - E)^2}{E}.$$

The χ^2 value for the class 650 to 700 is:

$$\chi^2 = \frac{(8 - 4.53)^2}{4.53} = 2.66$$

Step 10. Sum the individual χ^2 values to obtain the total χ^2 for this distribution. ($\chi^2 = 22.40$)

Step 11. Determine the significance by referring to a table of cumulative distribution of Chi-square (Appendix B). The number of degrees of freedom (d.f.) is the number of classes (k) minus the number of parameters estimated (p), minus one (d.f. = k-p-1). In this case d.f. = 18 – 2 – 1 = 15. Since χ^2 is less than the critical value for the 5% probability level with 15 degrees of freedom ($\chi^2 = 22.40 < 25.00$), the null hypothesis of a log normal distribution is not rejected, and the distribution is considered log normal.

BEST DOCUMENT AVAILABLE

Table 6. CDC Proficiency Testing Data for IgG

1 Class Limit	2 Class Midpoint	3 Log of Class Limit	4 Class Limit in Z	5 Area between O and Z	6 Area within Class (P)	7 Frequency Observed (O)	8 Frequency Expected (E)	9 $\chi^2 = \frac{(O-E)^2}{E}$	10 Observed Cumulative Frequency	11 Expected Cumulative Frequency	12 Deviation $\lvert D' \rvert$
≤ 500				−.5000							
	525				.0001	1	.03		1	.03	.97
550		6.3099	−3.70	−.4999							
	575				.0009	1 } 4	.27 } 1.60	3.60	2	.30	1.70
600		6.3969	−3.10	−.4990							
	625				.0044	2	1.30		4	1.60	2.40
650		6.4770	−2.55	−.4946							
	675				.0153	8	4.53	2.66	12	6.13	5.87
700		6.5511	−2.04	−.4793							
	725				.0375	5	11.10	3.35	17	17.23	0.23
750		6.6201	−1.57	−.4418							
	775				.0732	13	21.67	3.47	30	38.90	8.90
800		6.6846	−1.12	−.3686							
	825				.1075	27	31.82	0.73	57	70.72	13.72
850		6.7452	−0.71	−.2611							
	875				.1394	49	41.26	1.45	106	111.98	5.98
900		6.8024	−0.31	−.1217							
	925				.1456	44	43.10	0.02	150	155.08	5.08
950		6.8565	0.06	.0239							
	975				.1352	50	40.02	2.49	200	195.10	4.90
1000		6.9078	0.41	.1591							
	1025				.1143	33	33.83	0.02	233	228.93	4.07
1050		6.9565	0.75	.2734							
	1075				.0843	25	24.95	0.00	258	253.88	4.12
1100		7.0031	1.07	.3577							
	1125				.0554	17	16.40	0.02	275	270.28	4.72
1150		7.0475	1.36	.4131							
	1175				.0384	11	11.37	0.01	286	281.65	4.35
1200		7.0901	1.66	.4515							
	1225				.0223	3	6.60	1.96	289	288.25	0.75
1250		7.1309	1.94	.4738							
	1275				.0126	3	3.73	0.14	292	291.98	0.02
1300		7.1701	2.21	.4864							
	1325				.0068	1	2.01	0.51	293	293.99	0.99
1350		7.2079	2.47	.4932							
	1375				.0068	3	2.01	1.97	296	296.00	0.00
≥ 1400				.5000							
								$\chi^2 = 22.40$			$D' = 13.72$
			$Z = \frac{x - \bar{x}}{s}$								

2. The Kolmogorov-Smirnov (K-S) Test *(60,129)*

In most situations the K-S test is more powerful than the Chi-square test for determining the tenability of a hypothesized distribution. Let $F_0(x)$ be the completely specified cumulative distribution function based on n observations. For an observation x, $S_n(x) = K/n$ where K is the number of observations equal to or less than x. The largest value of $|F_0(x) - S_n(x)|$ is called the maximum deviation, D.

$$D = \text{Max } |F_0(x) - S_n(x)|$$

In other words, the type of frequency distribution is determined by comparing the theoretical distribution to be tested with the actual distribution observed. At points along the curve the difference between the curves is measured. The largest of these differences is then compared with the value that would be expected to result from random variation.

Under the null hypothesis we would expect the difference between $F_0(x)$ and $S_n(x)$ to be within the limits of random errors. For the chosen significance level, a value of D which is greater than or equal to the critical value specified for the K-S test would indicate that the null hypothesis should be rejected, i.e., that the theoretical and sample distributions are significantly different.

Let us use the sample example as in the Chi-square goodness of fit (Table 6). Columns 10 and 11 show the observed and expected cumulative frequencies, respectively. These were obtained by sequentially adding the frequencies in columns 7 and 8, respectively. Column 12 is simply the absolute value of the difference between the frequency in column 10 and that in column 11. The largest deviation (D′) is the 13.72 value observed in the 800 to 850 class. D is then calculated from the formula $D = D'/n$. In the example $D = 13.72/296 = 0.0464$, and the critical value of K-S test D, at the 5% probability level, with n = 18 determined from Appendix C equals 0.309. This is greater than the difference in this example ($D = 0.0464$), so, we cannot reject the null hypothesis. Therefore, the distribution of data is considered log-normal.

C. COMMON STATISTICAL TESTS

1. Determining Confidence Intervals

The best point estimate of a population mean (μ) is the sample mean ($\bar{x}$, $\bar{x}_G$ for geometric mean), and the best point estimate of a population standard deviation (σ) is the sample standard deviation (S, S_G for geometric standard deviation). Since point estimates represent only single value on the real number scale and provide no measure of how close we can expect them to be to the quantities they are supposed to estimate, they should always be accompanied by a confidence interval. Confidence intervals enable us to obtain useful information about population parameters without having to treat such parameters as statistical variables. It should be clearly understood that the level of confidence is merely a statement quantifying the percentage of times (for repetitions of the experiment) that intervals computed from this procedure would be expected to include the true value of the parameter (*73*).

As the Central Limit Theorem states "If n is large, the sampling distribution of a mean can be approximated closely with a normal distribution." The confidence intervals (C.I.) of a sample mean used to estimate the population mean can be expressed in the following form:

$$\text{C.I.} = \bar{x} \pm z \frac{\sigma}{\sqrt{N}} \text{ ; if } \sigma \text{ is known; otherwise}$$

$$= \bar{x} \pm t \frac{s}{\sqrt{N}} .$$

The z and t values are essentially equal when n is large (generally $n > 30$) and $z = t = 1.96$ with 95% confidence level. If $n < 30$, use a t table to determine the value of t for the desired sample size and confidence level.

As an example, the 95% confidence interval for a mean of 10 which was obtained from 50 individual tests and had a standard deviation of 3 would be:

$$95\% \text{ C.I.} = 10 \pm 1.96 \frac{3}{\sqrt{50}} = 10 \pm 0.83.$$

We can assert with 95% confidence that the true mean is between 9.17 and 10.83.

If the mean were derived from only ten tests then we must use the appropriate t value in place of the z value of 1.96. This is determined from Appendix D at the intersection of 9 degrees of freedom (N-1) and the probability level 5%. In this case the value is 2.262 and the 95% C.I. becomes:

$$95\% \text{ C.I.} = 10 \pm 2.26 \frac{3}{\sqrt{10}} = 10 \pm 2.15$$

With only ten values we can only assert with 95% confidence that the true mean is between 7.85 and 12.15.

To determine confidence intervals for geometric means the same formulas are used but the means and standard deviations must be converted to logarithms before calculating. For example, the 95% confidence interval for a geometric mean of 100 which was obtained from 50 individual tests and which had a geometric standard deviation of 2 would be:

$$95\%\ \text{C.I.} = \text{antilog}\left(\log 100 \pm 1.96\ \frac{\log 2}{\sqrt{50}}\right) = \text{antilog}\ (2 \pm 0.0834)$$

Therefore, we can assert with 95% confidence that the true geometric mean is between 82.5 (the antilog of 1.9165) and 121.2 (the antilog of 2.0834). This can conveniently be expressed as 100 x/÷ 1.212.

2. Determination of Sample Size

One question that always comes up when a study begins is "How many samples should I test?" In order to answer the question, we must know:

a. The desired degree of confidence

b. The population standard deviation (σ), or the best estimate of σ. For qualitative data, the expected probability of getting a positive result (p).

c. The size of estimated error (E) that can be tolerated. If the above information is known, we can write:

$$\text{Sample size } n = \left(\frac{z\sigma}{E}\right)^2 \quad \text{or} \quad \left(\frac{zs}{E}\right)^2$$

where z = 1.65 for 90% confidence level,

= 1.96 for 95% confidence level,

= 2.56 for 99% confidence level.

For example, if we want to be able to assert with a confidence of 95% that the estimate is not "off" by more than ± 2 units, and a reasonable estimate of σ equals 4 units, then

$$n = \left(\frac{1.96\ \ (4)}{2}\right)^2 = 15.37$$

and we round up to next nearest whole number. Hence a sample size of 16 will give the desired reliability within the specified allowable error.

For qualitative tests the sample size is determined by:

$$n = \left(\frac{Z\sqrt{pq}}{E} \right)^2$$

where p = probability of a positive result,
q = 1 − p = probability of a negative result.

Other symbols as in the quantitative example. E, p and q may be expressed either as proportions or as percentages, provided they are all expressed in the same units.

For example, if we want to be able to assert with a confidence of 95% that the estimate is not in error by more than 10 percentage points, and there is a 50% chance of a positive result, then

$$n = \left(\frac{1.96\sqrt{50 \times 50}}{10} \right)^2 = 96.04$$

A sample size of 97 will provide the desired results.

If the computed value of the sample size (n) is more than 10% of the population size (N), the finite population correction is needed. The revised sample size (n′) is obtained by

$$n' = \frac{n}{1 + \frac{n}{N}} .$$

For example, if in the qualitative example above the population size were 500, the revised n would be

$$n' = \frac{97}{1 + \frac{97}{500}} = 81.$$

3. Student's t Test

After the mean and standard deviation are estimated, a test can be performed to determine whether the theoretical and sample data are from the same population. In other words, is the difference between the estimate and its assumed value for the population statistically significant? The test statistic has a "t" distribution with n − 1 degrees of freedom.

$$t = \frac{\bar{x} - \mu_0}{S \big/ \sqrt{n}}$$

where: x = the sample mean,
μ_o = the population or theoretical mean,
S = the standard deviation,
n = number of samples tested.

For large samples drawn from a normally distributed population the distribution becomes very similar to the normal distribution. It can also be used for sampling from non-Gaussian distributions which have a well defined central tendency.

A two-tailed statistical test of a hypothesis concerning a population mean can be described as follows:

$$H_o: \mu = \mu_o \quad \text{vs} \quad H_1: \mu \neq \mu_o.$$

That is; the null hypothesis (H_o) states that the two populations are the same, and therefore the means will be equal. The alternative hypothesis (H_1) is that the means are not the same. Test statistic t is calculated from the formula above. If the calculated t is greater than the critical values of t, the null hypothesis will be rejected at the chosen level of significance and the means will be considered significantly different.

Example: An antistreptolysin O sample was sent to 325 laboratories around the nation. The results reported by laboratories were as follows:

Titer (y)	Log Titer (x)	Frequency (f)
120	4.78749	1
170	5.13580	1
240	5.48064	3
340	5.82895	10
480	6.17379	9
680	6.52209	41
960	6.86693	130
1360	7.21524	112
2720	7.90839	19
Total		326

$\bar{x} = 6.9285$

$S_x = .44997$

$S_{\bar{x}} = .02496$

$$\text{where } \bar{x} = \frac{\Sigma x}{n}, \quad S_x = \sqrt{\frac{\Sigma x^2 - \frac{(\Sigma x)^2}{n}}{n-1}} \quad \text{and } S_{\bar{x}} = \frac{S_x}{\sqrt{n}}$$

The t test is applied to determine if these results are in accord with a hypothesized value of 960 (log titer 6.86693). The null hypothesis is then that there is no difference between the mean for the sample and the theoretical value.

$H_0: \mu = 960$, and the alternate is that they are not the same; $H_1: \mu \neq 960$.

Since we know that serologic data have a log-normal distribution we will take the logarithm of each value and use the transformed data:

$\mu = 6.86693$
$\bar{x} = 6.9285$
$S_x = 0.44997$
$n = 326$

$$t = \frac{\bar{x} - \mu}{\frac{S_x}{\sqrt{n}}} = \frac{6.9285 - 6.8669}{\frac{.45}{\sqrt{326}}} = \frac{0.0616}{0.02491} = 2.47.$$

The critical value for t for a test with 325 (n−1) degrees of freedom at the 5% level is 1.96 (Appendix D). Since this is less than the calculated value for t (2.47), the null hypothesis is rejected and the conclusion is drawn that the average titer is not 960. There is less than a 5% chance that the titer (μ) is 960.

There are many situations in which we want to know whether there is a statistically significant difference between two sample means, e.g., when two methods are being compared. The t test can also be applied in this situation. We will discuss both two independent group t and paired t tests next. The paired t test is used when the results are paired, such as when two different tests are performed on each of a number of sera or when samples are matched in some way that is relevant to the test result. The two-group t test is used when there is not a basis for pairing. It is inappropriate to use the two group t test for paired data or vice versa.

a. **Two Group t Test:** When two samples come from similar populations, or from two normal populations with approximately equal variances, the test statistic is as follows:

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - D}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

where D = desired difference between two means (usually zero), and
S_p = Pooled standard deviation

BEST DOCUMENT AVAILABLE

$$Sp = \sqrt{\frac{(n_1 - 1)\,S_1^2 + (n_2 - 1)\,S_2^2}{n_1 + n_2 - 2}}.$$

Example: After results were reported for the first ASO samples indicated in example 1, the second ASO samples were sent to laboratories. The results were as follows:

Titer (y)	**Log Titer (x)**	**Frequency (f)**
60	4.90434	1
120	4.78769	1
170	5.13580	1
240	5.48064	4
340	5.82895	7
480	6.17379	10
680	6.52209	35
960	6.86693	127
1360	7.21524	124
2720	7.90839	16
Total		326

$\bar{x} = 6.9182$

$S_x = .49959$

$S_{\bar{x}} = .02810$

$$\text{where } \bar{x} = \frac{\Sigma x}{n},\quad S_G = \sqrt{\frac{\Sigma x - \frac{(\Sigma x)^2}{n}}{n-1}} \quad \text{and } S_{\bar{x}} = \frac{S_x}{\sqrt{n}}.$$

$H_0: \mu_1 = \mu_2$

$H_1: \mu_1 \neq \mu_2$

$$S_p = \sqrt{\frac{(326-1)\,(0.4497)^2 + (326-1)\,(0.49959)^2}{326 + 326 - 2}} = \sqrt{0.22603} = 0.47543$$

$$\text{Hence } t = \frac{(6.9285 - 6.9182) - 0}{0.47543\sqrt{\frac{1}{326} + \frac{1}{326}}} = \frac{0.0103}{0.03724} = 0.2766.$$

There are 650 degrees of freedom, and the critical value of t for a test at the 5% level is 1.96 (from Appendix D), which is greater than t = 0.2766. Hence, we cannot reject the null hypothesis and must conclude that the two samples are not statistically different at the 5% level of significance.

b. Paired t Test: For two samples whose observations are not obtained completely independently of one another, the paired t test is more powerful than the two-group t test. Hence, for paired observations a paired t test should be used to control for variation among observations from different pairs.

Example: A rubella sample was sent to ten laboratories which were asked to test it by two different methods. The reported results were as follows:

Lab	Titer with Heparin $MnCl_2$	x_1 = Log Titer	Titer with Kaolin	x_2 = Log Titer	$D = x_1 - x_2$
A	20	2.99573	10	2.30259	.69315
B	16	2.77259	8	2.07944	.69315
C	20	2.99573	20	2.99573	0
D	40	3.68888	10	2.30259	1.38629
E	32	3.46574	16	2.77259	.69315
F	16	2.77259	16	2.77259	0
G	16	2.77259	8	2.07944	.69315
H	32	3.46574	16	2.77259	.69315
I	40	3.68888	40	3.68888	0
J	20	2.99573	10	2.30259	.69315

The statistic for the paired t test is the same as for the usual t test, with the exception that the random variable x is replaced by the difference (D) between the two random variables, x_1, and x_2, and the standard error of the mean is replaced by the standard error of the difference, $\frac{S_D}{\sqrt{n}}$ Thus, for this example,

$$t = \frac{D}{\frac{S_D}{\sqrt{n}}} = \frac{.55452}{\frac{.43828}{\sqrt{10}}} = \frac{.55452}{.138628} = 4.00$$

where:

$$\bar{D} = \frac{\Sigma D}{n}, \quad S_D = \sqrt{\frac{\Sigma D^2 - \frac{(\Sigma D)^2}{n}}{n-1}}.$$

Since the calculated t value of 4.00 is greater than the critical value for t for a test at the 5% significance level with 9 degrees of freedom (2.252), we may conclude that these two methods are not the same.

4. Analysis of Variance

In the last section we discussed the comparison of two means. Now we will generalize the work and consider the problem of comparing more than two sample means. Before we proceed, the assumptions are restated: All samples are either drawn from normal populations or from approximately normally distributed populations with the same population standard deviation. The null hypothesis is that the means are equal.

The analysis of variance procedure is an attempt to analyze the sources of variation in responses and to estimate how much of the total variation each source causes. This permits the laboratorian to locate the important sources of variation in a study and to determine how they interact and affect the response (*58*).

The method involved in performing an analysis of variance can best be illustrated by considering an example. If the problem is to compare the IgG results using four different antibody sources, the null hypothesis will be that all means will be equal and that any difference among them can be attributed to chance.

Because we are comparing more than two means, neither the paired t nor two-group t test can be applied to such a situation. The results could be compared in a series of t tests on each of the possible pairs, but this procedure is likely to indicate significant differences when they do not exist.

The procedure is based on the observation that each of the pieces of the total sum of squares of deviations $(\Sigma(x - \bar{x})^2)$, divided by an appropriate constant, provides an independent and unbiased estimate of the variance of the experimental error. When a particular source of variation makes a major contribution to the error, its proportion of the error (called the "sum of squares" for the variable) will be large. This condition can be detected by comparing the estimate of variance for a particular independent variable with that obtained from the sum of squared deviations assigned to error, using an F test. If the estimate for the independent variable is significantly larger, the F test will reject the null hypothesis that the independent variable is not significantly larger than the others (*58*).

Example: Ten IgG samples were tested with four different antibodies to compare their results (Table 7). Are there any differences among the antibodies statistically?

Step 1. Since IgG results have a log-normal distribution, the first step is to convert all values by using a logarithmic transformation. In this case we used natural logarithms ($\log_e$ or ln).

Step 2. Three different kinds of sums of squares of deviations $(\Sigma(x - \bar{x})^2)$ are calculated: total sum of squares (TSS), sum of squares due to treatment (SST), and sum of squares due to error (SSE).

$$TSS = \Sigma\Sigma(x - \bar{x})^2 = \Sigma\Sigma x^2 - \frac{(\Sigma\Sigma x)^2}{kn}$$

$$SST = n\Sigma(\bar{x}_i - \bar{x})^2 = n\left(\Sigma\bar{x}_i^{\,2} - \frac{(\Sigma\bar{x}_i)^2}{k}\right)$$

$$SSE = \Sigma\Sigma(x - \bar{x}^2)^2 = TSS - SST$$

where k = number of groups and,
n = number in each group,
$\bar{x}_i$ = mean for each group,
$\bar{x}$ = grand mean.

In this example:

$$TSS = 1911.32472 - \frac{(276.48573)^2}{4 \times 10}$$

$$= 1911.32472 - 1911.108972$$

$$= 0.215748$$

$$SST = 10\left(191.11426 - \frac{(27.64858)^2}{4}\right)$$

$$= 0.03266$$

$$SSE = 0.18312$$

Step 3. Compute mean square for treatment (MST) and error (MSE).
$MST = SST/(k - 1) = 0.03266/3 = 0.01089$
$MSE = SSE/[k(n - 1)] = 0.18312/36 = 0.00509$.

Step 4. Compute F value to test null hypothesis
$F = MST/MSE \cdot 0.01089/0.00509 = 2.14$.

Step 5. The critical value of the F statistic is approximately 2.86 at 5% level (Appendix E), which is greater than 2.14. Hence we cannot reject the null hypothesis and conclude that there are no differences among four antibody sources.

In this section we have only discussed the most simple one-way analysis of variance. For more complicated design refer to any statistical text.

5. Components of Variance Analysis

Because variance (S^2) has the property of additivity, it is useful in determining how much each source of error contributes to the overall variation in test results. With this information and

Table 7. IgG Results Compared by Antibody Source

IgG results on 10 samples were compared by antibody source with the following results:

	Ab A		Ab B		Ab C		Ab D		
Sample No.	mg/dl	Log	mg/dl	Log	mg/dl	Log	mg/dl	Log	Total Values
1	993	6.90073	1000	6.90776	995	6.90274	1110	7.01212	
2	820	6.70930	1116	7.01751	1030	6.93731	998	6.90575	
3	943	6.84907	1090	6.99393	1134	7.03351	1000	6.90776	
4	951	6.85751	996	6.90375	899	6.80128	1060	6.96602	
5	1000	6.90776	976	6.88346	1010	6.92658	926	6.83087	
6	1110	7.01212	954	6.86066	1000	6.90776	1087	6.99118	
7	876	6.77537	946	6.85224	926	6.83087	1000	6.90776	
8	1010	6.91771	980	6.88755	1034	6.94119	1025	6.93245	
9	980	6.88755	1050	6.95655	1065	6.97073	975	6.88244	
10	920	6.82437	1200	7.09008	1010	6.91771	1080	6.98472	
Σx		68.64149		69.35349		69.16968		69.32107	$\Sigma\Sigma x$ 276.48573
Σx^2		471.22768		481.04445		478.48311		480.56948	$\Sigma\Sigma x^2$ 1911.32472
$\bar{x}_i$	957	6.86415	1028	6.93535	1009	6.91697	1025	6.93211	$\Sigma\bar{x}_i$ = 27.64858
$\bar{x}_i^2$		47.11656		48.09908		47.84447		48.05415	$\Sigma\bar{x}_i^2$ = 191.11426
									$\bar{x}$ 6.91214

knowledge of external constraints, the laboratorian can plan the most efficient ways to reduce error in the test results.

Additivity of variance means that the variance from each source of error (if known) can be added to obtain the total variance. Conversely, if one source of error can be eliminated, its contribution to the total variance can be determined by subtraction. For example, if a number of technologists perform replicate assays on the same sample, the amount of error contributed by each technologist can be determined by subtracting the within-technologist variance from the total variance. It is helpful to express this component of variance as a percentage of the total variance.

For arithmetic data, the variances can be added and subtracted directly, but for geometric data the variance must be in the log form. The geometric standard deviation, S_G, cannot be converted to a variance by squaring it, but the log of the standard deviation must be squared. Logarithms to any base may be used, but the values are different for different bases. For the sake of convenience and uniformity it has been recommended that log to the base 2 be used. Variance in this form is obtained by dividing the log of the geometric standard deviation by log 2 and squaring the result.

$$\text{Variance}_{\log 2} = \frac{\log_{10} S_G{}^2}{\log_{10} 2} \text{ or } \frac{\ln S_G{}^2}{\ln 2}$$

$$= \frac{\log_{10} S_G{}^2}{0.301} \text{ or } \frac{\ln S_G{}^2}{0.693}.$$

To obtain the S_G from the variance to the log base 2 use:

$$S_G = \text{antilog}_{10}(\sqrt{\text{variance}} \times \log_{10} 2) \text{ or antiln } (\sqrt{\text{variance}} \times \ln 2)$$

$$= \text{antilog}_{10}(\sqrt{\text{variance}} \times 0.301) \text{ or antiln } (\sqrt{\text{variance}} \times 0.693).$$

An example using this technique is shown in a rheumatoid factor standardization study (*36*). Before standardization, the geometric standard deviation was 4.15 and the variance was 4.21, but after standardization the geometric standard deviation was 1.98 and the variance was 0.97. Therefore, of the total variance 3.24 or 77% was the result of lack of standardization.

6. Nonparametric Tests (*96*)

Most of the methods in the previous sections required specific assumptions about the population(s). They must be either normal or approximately normal. The observations must be independent and must be measured on at least an interval scale. Since there are many situations in which some or all of these requirements cannot be met, nonparametric tests provide alternate techniques based on less stringent assumptions. Nonparametric tests are very useful when applied to data that are difficult to quantify and are particularly useful in making inferences when serious doubt exists about the distribution of results.

a. **Sign Test:** The sign test uses plus and minus signs rather than quantitative measures to study two related samples.

The null hypothesis is that the two population distributions are identical, and we would expect the number of pair-positive differences to be equal to the number of pair-negative differences.

Example: Two rubella samples were sent to each of ten laboratories for analysis. The titers were reported as follows:

Lab	Sample I	Sample II	Sign*
A	32	64	–
B	40	20	+
C	128	128	0
D	64	128	–
E	64	32	+
F	80	40	+
G	20	80	–
H	32	128	–
I	64	128	–
J	64	256	–

*+ means sample I has higher titer than sample II
– means sample II has higher titer than sample I, and
0 implies that it is not possible to discriminate between the results of sample I and sample II

We count the number of signs in the smaller group ($x = 3$), and then the number of matched pairs which showed differences ($N = 9$). The probability computed for the sign test (two-tailed) is a binomial expression:

$$p = 2\sum \binom{N}{x} p^x q^{N-x} = 2\sum \binom{N}{x} (0.5)^N = \frac{2\sum \binom{N}{x}}{2^N}$$

where $\binom{N}{x}$ indicates the number of combinations of N things taken x at a time. This expression can be evaluated by the expression

$$\binom{N}{x} = \frac{N!}{x!\,(N-x)!}$$

x! (x factorial) means the product of all the numbers from x to 1. For example, 5! = 5 x 4 x 3 x 2 x 1. 0! is defined as equal to 1.

In this example, we want to know the probability of getting 3 or fewer signs; therefore

$$p = \frac{2\Sigma\binom{N}{x}}{2^N} \text{ for } x = 0, 1, 2, \text{ and } 3.$$

$$p = 2\left[\frac{\binom{9}{0} + \binom{9}{1} + \binom{9}{2} + \binom{9}{3}}{2^9}\right]$$

$$= 2\left[\frac{\frac{9!}{0!\,(9-0)!} + \frac{9!}{1!\,(9-1)!} + \frac{9!}{2!\,(9-2)!} + \frac{9!}{3!\,(9-3)!}}{512}\right]$$

$$= 2\left[\frac{1 + 9 + 36 + 84}{512}\right]$$

$$= 2\ \left[\frac{130}{512}\right] = 0.51$$

For the two-tailed test we have determined that the probability of occurrence by chance is .51 which is greater than the traditionally used 5% significance level; thus we cannot reject the null hypothesis and must conclude that the two rubella samples do not have significantly different titers.

For a one-tailed test the procedure is the same but the formula is:

$$p = \Sigma\binom{N}{x}p^x\, q^{N-x} = \frac{\Sigma\binom{N}{x}}{2^N}.$$

b. **Fisher Exact Probability Test:** For discrete data, the Fisher exact test is very useful in analyzing two small independent samples. The test determines whether the two classes (I and II) differ in the proportions that are in two classifications (+ and –).

	+	-	Total
I	A	B	A+B
II	C	D	C+D
Total	A+C	B+D	N

We know, when the marginal totals are regarded as fixed, that the probability that the observed results occurred by chance is:

$$p=\frac{\binom{(A+C)}{A}\binom{(B+D)}{B}}{\binom{N}{(A+B)}}$$

which can be evaluated by the equation

$$\binom{N}{x}=\frac{N!}{x!\,(N-x)!}$$

$$p=\frac{\binom{(A+C)}{A}\binom{(B+D)}{B}}{\binom{N}{(A+B)}}=\frac{\frac{(A+C)!}{A!C!}\times\frac{(B+D)!}{B!D!}}{\frac{N!}{(A+B)!\,(N-(A+B))!}}=\frac{\frac{(A+C)!(B+D)!}{A!\,B!\,C!\,D!}}{\frac{N!}{(A+B)!(C+D)!}}$$

$$=\frac{(A+B)!\,(C+D)!\,(A+C)!\,(B+D)!}{N!A!B!C!D!}$$

To illustrate the use of this formula, suppose 27 persons were examined for a disease with two different tests, and the results were as shown below:

	Result		
Test	+	-	Total
I	2	12	14
II	7	6	13
Total	9	18	27

The probability of obtaining the results shown by chance alone is:

$$p_2=\frac{14!13!9!18!}{27!2!12!7!6!}=\frac{364}{10295}=0.03332.$$

To determine the probability of obtaining the given results or results showing a greater difference between classes we reduce the smallest cell by one and keep the marginal totals fixed until the smallest cell is zero.

	Result		
Test	+	−	Total
I	1	13	14
II	8	5	13
Total	9	18	27

$$p_1 = \frac{(1+13)!(8+5)!(1+8)!(13+5)!}{27!1!13!8!5!} = \frac{14!13!9!18!}{27!1!13!8!5!}$$

$$= \frac{42}{10925} = 0.00284$$

	Result		
Test	+	−	Total
I	0	14	14
II	9	4	13
Total	9	18	27

$$p_0 = \frac{(0+14)!(9+4)!(0+9)!(14+4)!}{27!0!9!14!4!} = \frac{14!13!9!18!}{27!0!9!14!4!} = 0.00015.$$

Thus the probability of obtaining 2 or fewer positives by test I by chance alone is:

$$p = p_0 + p_1 + p_2 = 0.03332 + 0.00384 + 0.00015 = 0.03731.$$

For use of 5% probability, we reject the null hypothesis because 0.03731 is less than 0.05. We conclude that significantly more positive results were reported with test II than with test I.

The Fisher exact test should always be used when the sample size is less than 20 and/or when any expected frequencies are less than 5. Otherwise, Chi-square test is more convenient to use.

c. **Chi-Square Test:** The Chi-square test is very useful in determining the significance of differences between two independent classes. For the 2 x 2 Chi-square test when N is between 20 and 40, all expected frequencies must be 5 or more, otherwise, the Fisher exact test is recommended.

Let us use the same example as for the Fisher exact test, with the following observed frequencies:

	Result		
Test	+	−	Total
I	5	20	25
II	10	10	20
Total	15	30	45

$$\chi^2 = \frac{N(\,|AD-BC| - \frac{N}{2})^2}{(A+B)(C+D)(A+C)(B+D)}$$

$$= \frac{45(\,|(5 \times 10) - (20 \times 10)| - \frac{45}{2})^2}{25 \times 20 \times 15 \times 30} = 3.25.$$

d.f. = (r − 1) x (c − 1) where r = number of rows and c = number of columns.

Since the critical value of Chi-square with 1 degree of freedom is 3.84 (Appendix B), which is greater than 3.25, the null hypothesis is not rejected at 5% level, and we conclude that there is no significant difference between tests A and B.

In the formula above the $\frac{N}{2}$ correction term for continuity was suggested by Yates to bring the probability associated with χ^2 into close agreement with the more exact probabilities.

For tests with more than two independent samples, this correction is not needed; refer to Siegel (96).

We have discussed some of the most frequently used parametric and nonparametric statistical tests. The following table gives a summary of the tests and calculations discussed and references to the discussion in this monograph and other sources.

BLANK PAGE

Table 8. Appropriate Tests and Calculations for Statistical Analyses

Desired Analysis or Determination	Test Required	References
Determine the shape of a frequency distribution (test the "Goodness of Fit" of a sample distribution to a theoretical distribution).	Chi-Square (χ^2) or	Sect. VI, B,1
	K-S	Sect. VI, B,2 (60)
Determine distribution parameters (estimate population mean and standard deviation).		Sect. VI, A
Normal distribution	(Arithmetic) $\bar{x}$ and S	(19,58,73,75,100, 114,121)
Log-normal distribution	(Geometric) $\bar{x}_G$ and S_G	Sect. VI,A (112)
Determine Confidence Interval	C.I.	Sect. VI,C,1
Determine required sample size		Sect. VI,C,2
Determine if sample data and theoretical data differ:		
If sample data significantly higher (or lower) than theoretical data?	One-tailed t test	
Is sample data significantly higher or lower than theoretical data?	Two-tailed t test	Sect. VI,C,3
Are two sample distributions significantly different?		
Two independent groups:		
Normal distribution or transformable to normal distribution.	Two-group t test (independent t test)	Sect. VI,C,3a
Any distribution of discrete data.		
Small sample size ($n < 20$ and/or $f < 5$)*	Fisher exact probability test	Sect. VI,C,6b (96)
Large sample size	Chi-square (χ^2) test	Sect. VI,C,6c
Paired data:		
Normal distribution or transformable to normal distribution	Paired t test (dependent t test)	Sect. VI,C,3b
Any distribution	Sign test	Sect. VI,C,6a (96)
Are there any significant differences among more than two groups?	Analysis of Variance (F test)	Sect. VI,C,4
How much does each source of error contribute to the total error?	Components of Variance Analysis	Sect. VI,C,5
Establishment of a Quality Control Chart		
Normal data	Arithmetic QC Chart	Sect. VI,C,1 (56,114,121)
Log-normal data	Geometric QC Chart	Sect. VI,A (112)
To change logarithms to a different base	$\text{Log}_a \rightarrow \text{Log}_b$	Table 5

*n = total number of samples
f = frequency in smallest cell

BLANK PAGE

BIBLIOGRAPHY

1. Anderson S.G. and E.J. Addison. 1971. Antinuclear-factor serum (homogeneous): an international collaborative study of the proposed research standard 66/233. *In* Beutner, E.H. (ed.). Defined Immunofluorescent Staining. Ann. N.Y. Acad. Sci. 177:337.

2. Anderson, S.G. and E.J. Addison. 1970. Progress towards immunofluorescence standards. *In* Holborow, E.J. (ed.). Standardization in Immunofluorescence. Oxford, Blackwell Scientific Publications Part 1, p. 1.

3. Anderson, S.G., D.R. Baugham, I. Batty, W. Becker, B. Cinadar, M. van der Giessen, F. Hymes, D. Long, F. Peetoom, K. Pondman, M. Raynand, C.B. Reimer, D.S. Rowe, H.G. Schwick and S. Vrancheva. 1971. Measurements of the concentrations of human serum immunoglobulins. J. Immunol. 107:1798-1799.

4. Anderson, S.G., N.W. Bentzva, and V. Houba. 1970. International reference preparation of rheumatoid arthritis serum. Bull. WHO. 42:311-318.

5. Bailey, G.H. and S. Raffel. 1935. Hemolytic antibodies for sheep and ox erythrocytes in infectious mononucleosis. J. Clin. Invest. 14:228-244.

6. Bartlett, R.C. 1974. Medical Microbiology: Quality, Cost, and Clinical Relevance. Wiley-Interscience, New York, NY.

7. Bartlett, R.C. 1975. Functional quality control. p. 145-149. *In* Prier, J.E., J.T. Bartola, and H. Friedman (eds.). Quality Control in Microbiology. University Park Press, Baltimore, MD.

8. Batty I. and G. Torrigrani. 1976. Standardization of reagents and methodology in immunology, p. 911-917. *In* Rose, N.R. and H. Friedman (ed.). Manual of Clinical Immunology. American Society for Microbiology, Washington D.C.

9. Belk, W.P. and F.W. Sunderman. 1947. A survey of the accuracy of chemical analyses in clinical laboratories. Am. J. Clin. Path. 17:853-861.

10. Berson, S.A. and R.S. Yalow. 1958. Assay of plasma insulin in human subjects by immunologic methods. Nature 184:1648-1649.

11. Berson, S.A. and R.S. Yalow. 1964. Immunoassay of the protein hormones, p. 557. *In* Pincus G., K.V. Thimann and E.D. Astwood (eds.). The Hormones. Academic Press, New York, NY.

12. Berson, S.A. and R.S. Yalow. 1968. General principles of radioimmunoassay. Clin. Chem. Acta 22:51.

13. Biano, S.A., C.C. Jenkins and J.B. Daniels. 1971. Effect of calcium concentration on inhibitors of hemagglutination of rubella virus. J. Inf. Diseases 123:421-425.

14. Biological Products Division. 1974. Specifications and Evaluation Methods for Immunological and Microbiological Reagents. II. Chlamydial, Mycoplasmal, Rickettsial, and Viral (4th ed.). Public Health Service, Center for Disease Control, Atlanta, GA.

15. Biological Products Division. 1975. Specifications and Evaluation Methods for Immunological and Microbiological Reagents. I. Bacterial, Fungal, and Parasitic. (4th ed.). Public Health Service, Center for Disease Control, Atlanta, GA.

16. Biological Products Division. 1977. Catalog of Immunological and Microbiological Products. Center for Disease Control, Atlanta, GA.

17. Bodily, H.L. 1970. General administration of the laboratory. p. 1-28. *In* Bodily, H.L., E.L. Updyke and J.O. Mason (ed.). Diagnostic Procedures for Bacterial, Mycotic and Parasitic Infections (5th ed.). American Public Health Association, Inc. New York, NY.

18. Boone, D.J., G.O. Guerrant and R.W. Knouse. 1977. Proficiency testing in clinical toxicology: Program sponsored by the Center for Disease Control. J. Anal. Toxicol. 1:147-150.

19. Boutwell, J.H. 1970. Quality Control – Clinical Chemistry. Center for Disease Control, Atlanta, GA.

20. Bureau of Laboratories. 1977. Recommended tests for the serologic detection of infection by group A hemolytic streptococci. Current Item #258. U.S. Department of H.E.W., Center for Disease Control, Atlanta, GA.

21. Bureau of Laboratories. 1978. Infectious Mononucleosis Serology. Current Item #259. U.S. Department of H.E.W., Center for Disease Control, Atlanta, GA.

22. Casey, H.L. 1969. Importance of control mechanisms in diagnostic serology. *In* Balows, A. and H.D. Isenberg (eds.). Quality Control in the Clinical Microbiology Laboratory. A Roundtable Discussion. General Diagnostics Division of Warner-Lambert Pharmaceutical Co.

23. Cavallaro, J.J., D.F. Palmer and P.E. Bigazzi. 1976. Immunofluorescence Detection of Autoimmune Diseases. Center for Disease Control, Public Health Service, U.S. Department of H.E.W., Atlanta, GA.

24. Center for Disease Control. 1977. Rubella Hemagglutination – Inhibition Tests. A Procedural Guide. Rev. 8-77. Center for Disease Control, Atlanta, GA.

25. Cherry W.B., M. Goldman, T.R. Carski and M.D. Moody. 1960. Fluorescent Antibody Techniques in the Diagnosis of Communicable Diseases. Public Health Service Publication #729. Communicable Disease Center, Atlanta, GA.

26. Cherry, W.B. and C.B. Reimer. 1973. Standardization of diagnostic materials. 4. Diagnostic immunofluorescence. Bull. WHO 48:737-746.

27. Copeland, B.E. 1974. Statistical tools in clinical pathology. p. 1-14. *In* Davidsohn, I. and J.B. Henry (eds.). Clinical Diagnosis by Laboratory Method. (15th ed.). W.B. Saunders Co., Philadelphia, PA.

28. Davis, N.E. and M. Ho. 1976. Quantitation of Immunoglobulins. p. 4-16. *In* Rose, N.R. and H. Friedman (eds.). Manual of Clinical Immunology. American Society for Microbiology, Washington D.C.

29. Deodhar, S.D., W.F. Braun, L.P. Cawley, P.W. Keitges, R.M. Nakamura, G.M. Penn, G. Reynoso and E.S. Tucker III. 1978. Immunology. p. 747-786. *In* Inhorn, S.L. (ed.). Quality Assurance Practices for Health Laboratories. American Public Health Association, Washington, D.C.

30. Ekins, R.P., G.B. Newman and J.L.H. O'Riordan. 1968. Theoretical aspects of "saturation" and radioimmunoassay. p. 59. *In* Hayes, R.L., F.A. Goswitz and B.E.P. Murphey (eds.). Radioisotopes in Medicine: U.S. Atomic Energy Commission, Oak Ridge, TN.

31. Ellis, R.J. 1976. Manual of Quality Control Procedures for Microbiological Laboratories. Center for Disease Control, Atlanta, GA.

32. Elveback, L.R., C.L. Guillier and F.R. Keating, Jr. 1970. Health, normality, and the ghost of Gauss. J.A.M.A. 211:69-75.

33. Engvall, E., K. Jonsson and P. Perlmann. 1971. Enzyme-linked immunosorbent assay II. Quantitative assay of protein antigen, immunoglobulin G, by means of enzyme-labelled antigen and antibody coated tubes. Biochim. Biophys. Acta 251:427-434.

34. Engvall, E., and H.E. Carlsson. 1976. Enzyme-linked immunosorbent assay, ELISA. p. 135-147. *In* Feldman, G., P. Druet, J. Bignon and S. Avrameas. (eds.). Immunoenzymatic Techniques. North Holland Publishing Co. Amsterdam.

35. Faulk, W.P. and W. Hijmans. 1972. Recent developments in immunofluorescence. Prog. Allergy 16:9-39.

36. Fulford, K.M., R.N. Taylor and V.A. Przybyszewski. 1978. Reference preparation to standardize results of serological tests for rheumatoid factor. J. Clin. Micro. 7:434-441.

37. Galen, R.S. and S.R. Gambino. 1975. Beyond Normality: The Predictive Value and Efficiency of Medical Diagnosis. John Wiley & Sons, New York, NY.

38. Garratty, G. and L.D. Petz. 1971. An evaluation of commercial antiglobulin sera with particular reference to their anticomplement properties. Transfusion. 11: 79-88.

39. Garvey, J.S., N.E. Cremer and D.H. Sussdorf. 1977. Methods in Immunology. W.A. Benjamin, Inc., Reading, MA.

40. Gebott, M. 1976. The art of electrophoresis quality control. Beckman Electrophoresis Information 3:2-7.

41. Gilbert, R.K. 1975. Survey Data 1947. College of American Pathologists. Chicago, IL.

42. Goldman, M. 1968. Fluorescent Antibody Methods. Academic Press. New York, NY.

43. Goldsmith, S.J. 1975. Radioimmunoassay: Review of basic principles. *In* Freeman, L.M. and M.D. Blaufox (eds.). Radioimmunoassay. p. 1-28. Grune and Stratton, New York, NY.

44. Guerrant, G.O. and C.T. Hall. 1977. Drug abuse proficiency testing. Clin. Toxicol. 10:209-219.

45. Harrell, W.K. 1970. Serological procedures. p. 45-54. *In* Bodily, H.L., E.L. Updyke and J.O. Mason (eds.). Diagnostic Procedures for Bacterial, Mycotic and Parasitic Infections (5th ed.). American Public Health Association, Inc., New York, NY.

46. Hebert, G.A., B. Pittman, R.M. McKinney and W.B. Cherry. 1972. The Preparation and Physiochemical Characterization of Fluorescent Antibody Reagents. Center for Disease Control, Atlanta, GA.

47. Herner, A.E. and J.A. Clayton-Hopkins. 1978. Troubleshooting radioligand assays. Clin. Chem. 23:1275-1280.

48. Hijmans, W. and M. Schaeffer. 1975. The 5th International Conference on Immunofluorescence and the Related Staining Techniques. Ann. N.Y. Acad. Sci. 254:21-172.

49. Humphrey, J.H. and I. Batty. 1974. Internal reference preparations for human serum IgG, IgA and IgM. Clinical and Experimental Immunology 17:708.

50. Kawamura, A., Jr. 1969. Fluorescent Antibody Techniques and their Applications. University Park Press, Baltimore, MD.

51. Keitges, P.W. 1976. Proficiency testing and clinical laboratory immunology, p. 918-920. *In* Rose, N.R. and H. Friedman (eds.). Manual of Clinical Immunology. American Society for Microbiology, Washington D.C.

52. Klein, G.C. 1976. Immune response to streptococcal infection. p. 264-273. *In* Rose, N.R. and H. Friedman (eds.). Manual of Clinical Immunology. American Society for Microbiology, Washington D.C.

53. Klein, G.C., M.D. Moody, C.N. Baker and B.V. Addison. 1968. Antistreptolysin O test. Appl. Micro. 16:184.

54. Laurence, D.J. 1975. First British standard for carcinoembryonic antigen (CEA). Brit. J. Cancer 32:295-299.

55. Lennette, E.H. and N.J. Schmidt. 1973. Principles and Performance of the Rubella Hemagglutination Inhibition Test. American Society of Clinical Pathologists. Chicago, IL.

56. Levey, S. and E.R. Jennings. 1950. The use of control charts in the clinical laboratory. Am. J. Clin. Path. 20:1059-1066.

57. Martinek, R.G. 1978. Quality control based on patient's data. J. Am. Med. Tech. 40:55-60.

58. Mandel, J. 1964. The Statistical Analysis of Experimental Data. Interscience-Wiley, New York, NY.

59. Mason, J.K. 1961. An ox cell hemolysin test for the diagnosis of infectious mononucleosis. J. Hyg. 49:471-479.

60. Massey, J.F., Jr. 1951. The Kolmogorov-Smirnov test for goodness of fit. J. Am. Stat. Assn. 46:68-78.

61. McFarland, J. 1907. The nephelometer: An instrument for estimating the number of bacteria in suspensions used for calculating the opsonic index and for vaccines. J.A.M.A. 49:1176-1178.

62. McKinney, R.M., J.T. Spillane and G.W. Pearee. 1964. Determination of purity of fluorescein isothiocyanates. Analyt. Biochem 7:74-86.

63. Medical Research Council Working Party. 1971. Recommendations on the characterization of antisera as reagents. Immunology 20:3-10.

64. Millipore Corporation. Electrophoresis and Immunoelectrophoresis in the Clinical Laboratory. Application Report AR-22. Bedford, MA.

65. Morgan, W.J.B., I. Davidson and C.N. Herbert. 1973. The use of the second international standard for anti-*Brucella abortus* serum in the complement-fixation test. J. Biol. Stand. 1:43-61.

66. Nairn, R.C. (ed.). 1969. Fluorescent Protein Tracing (3rd ed.). Williams & Wilkins, Baltimore, MD.

67. Nakamura, R.M. 1974. Immunopathology: Clinical Laboratory Concepts and Methods. Little Brown and Company. Boston, MA.

68. National Animal Disease Laboratory (NADL) Diagnostic Reagents Manual 65D: Standard Agglutination Test Procedure for the Diagnosis of Brucellosis. Agricultural Research Service, U.S. Department of Agriculture, Ames, IA.

69. National Bureau of Standards. Catalog of National Bureau of Standards, Standard Reference Materials 1975-1976 Edition. 1975. National Bureau of Standards Special Publication #260. U.S. Department of Commerce, Office of Standard Reference Materials, Institute for Materials Research, National Bureau of Standards, Washington, D.C.

70. National Committee for Clinical Laboratory Standards. 1978. List of Standards. Villanova, PA.

71. National Communicable Disease Center. 1969. Manual of Tests for Syphilis. Public Health Service Publication #411. U.S. Department of H.E.W., Public Health Service, Center for Disease Control, Atlanta, GA.

72. Nelson, D.B., E.P. Quirin and S.L. Inhorn. 1974. Compatibility of trypsin – modified human erythrocytes in the rubella hemagglutination-inhibition test employing three serum treatment procedures. Applied Micro. 27:767-770.

73. Ostle, B. 1963. Statistics in Research. Iowa State University Press. Ames, IA.

74. Palmer, D.F., H.L. Casey, J.R. Olsen, V.H. Eller and J.M. Fuller, 1974. A Guide to the Performance of the Standardized Diagnostic Complement Fixation Method and Adaptation to Micro Test. U.S. Department of H.E.W., Public Health Service, Center for Disease Control, Atlanta, GA.

75. Palmer, D.F. and J.J. Cavallaro. 1976. Some concepts of quality control in immunoserology, p. 906-910. *In* Rose, N.R. and H. Friedman (eds.). Manual of Clinical Immunology. American Society for Microbiology, Washington, D.C.

76. Palmer, D.F., J.J. Cavallaro and R.H. Galt. 1973. Laboratory Diagnosis by Serologic Method. Course number 9440-C. Center for Disease Control, Public Health Service, U.S. Department of H.E.W., Atlanta, GA.

77. Palmer, D.F., K.L. Herrmann and J.J. Cavallaro. 1974. Serodiagnosis of Toxoplasmosis, Rubella, Cytomegalic Inclusion Disease, and Herpes Simplex. Center for Disease Control, Public Health Service, U.S. Department of H.E.W. Atlanta, GA.

78. Palmer, D.F., K.L. Herrmann, R.E. Lincoln, M.V. Hearn and T.M. Fuller. 1970. Standardized Rubella Hemagglutination – Inhibition test. Immunology series No. 2, Procedural Guide, U.S. Department of H.E.W., Atlanta, GA.

79. Palmer, D.F. and R. Woods. 1972. A Procedural Guide to the Performance of the Radial Immunodiffusion Test and the Immunoelectrophoresis Test for Qualitation and Quantitation of Immunoglobulins. Center for Disease Control, Atlanta, GA.

80. Peddecord, K.M. 1978. Clinical Laboratory Proficiency Test Performance: Its Relationship to Environmental Structural and Process Variables. Doctoral Dissertation. University of Texas, Houston, TX.

81. Perkins, F.T., F.W. Sheffield, A.S. Outschoorn and D.A. Hemsley. 1973. An international collaborative study on the measurement of the opacity of bacterial suspensions. J. Biol. Stand. 1:1-10.

82. Przybyszewski, V.A. and K.A. Kelley. 1979. Rubella serology: The effect of aging on the removal of nonspecific inhibitors of hemagglutination from lyophilized sera. (submitted for publication).

83. Rantz, L.A. and E. Randall. 1945. A modification of the technic for determination of the antistreptolysin titer. Proc. Soc. Exp. Biol. Med. 59:22-25.

84. Reimer, C.B. and S.F. Maddison. 1976. Standardization of human immunoglobulin quantitation: a review of current status and problems. Clin. Chem. 22: 577-582.

85. Rodbard, D. 1974. Statistical quality control and routine data processing for radioimmunoassay and immunoradiometric assays. Clin. Chem. 20:1255-1270.

86. Rodbard D., P.L. Rayford, J.A. Cooper and G.T. Ross. 1968. Statistical quality control of radioimmunoassays. J. Clin. Endocrinol. Metab. 28:1412-1418.

87. Rowe, D.S. 1971. Introductory review: characteristics of sera and conjugates. *In* Beutner, E.D. (ed.). Defined Immunofluorescent Staining. Ann. N.Y. Acad. Sci. 177:335-336.

88. Rowe, D.S., S.G. Anderson and L. Tackett. 1970. A research standard for human serum immunoglobulins D. Bull. W.H.O. 43:607-609.

89. Rowe, D.S., S.G. Anderson and B. Grab. 1970. A research standard for human serum immunoglobulins IgG, IgA and IgM. Bull. W.H.O. 42:535-552.

90. Rowe, D.S., B. Grab and S.G. Anderson. 1972. An international reference preparation for human serum immunoglobulins G, A and M: content of immunoglobulins by weight. Bull. W.H.O. 46:67-79.

91. Rowe, D.S., B. Grab and S.G. Anderson. 1973. An international reference preparation for human serum immunoglobulin E. Bull. W.H.O. 49:320-321.

92. Russell, R.L. 1974. Quality control in the microbiology laboratory, p. 862-870. *In* Lennette, E.H., E.H. Spaulding and J.P. Truant (eds.). Manual of Clinical Microbiology, Washington, D.C.

93. Satz, J.E. 1975. Quality control in the manual serological tests for syphilis. p. 105-118. *In* Prier, J.E., J.T. Bartola and H. Friedman (eds.). Quality Control in Microbiology. University Park Press, Baltimore, MD.

94. Schuurs, A.H.W.M. and B.K. Van Weemen. 1977. Enzyme-immunoassay. Clin. Chim. Acta 81:1-40.

95. Shortridge, K.F. 1977. Rubella virus serology: Detection of residual lipoprotein inhibitors of hemagglutination using sensitive indicator arboviruses. J. Clin. Path. 30:409-416.

96. Siegel, S. 1956. Nonparametric Statistics for the Behavioral Sciences. McGraw Hill Book Co. New York, NY.

97. Sizaret, P. and S.G. Anderson. 1976. The international reference preparation for alpha-fetoprotein. J. Biol. Stand. 4:149.

98. Sizaret, P. N. Breslow, S.G. Anderson and twelve other participants. 1975. Collaborative study of a preparation of human cord serum for its use as a reference in the assay of alpha-fetoprotein. J. Biol. Stand. 3:201-223.

99. Skelly, D.S., L.P. Brown and P.K. Besch. 1973. Radioimmunoassay. Clin. Chem. 19:146-186.

100. Snedecor, G.W. and W.G. Cochran. 1967. Statistical Methods (6th Ed). Iowa State University Press. Ames, IA.

101. Sokal, R.R. and F.J. Rohlf. 1969. Biometry. W.H. Freeman and Co. San Francisco, CA.

102. Spaun, J., M.M. Bentzon, S. Olsen Larsen and L.F. Hewitt. 1961. International standard for antistreptolysin O. Bull. W.H.O. 24:271-279.

103. Stein, S.F., J.H. Wave and R. Woods. 1977. Serum immunoglobulins: methods for the determination of normal values in international units. J. Immunol. Methods. 16:371-384.

104. Suggs, M.T. 1975. Product class standards (specifications) and evaluation of microbiological in vitro diagnostic reagents p. 87-104. *In* Prier, J.E., J.T. Bartola and H. Friedman (eds.). Quality Control in Microbiology. University Park Press, Baltimore, MD.

105. Taylor, R.N. 1974. Derivation of Evaluation Survey Data from Proficiency Testing for Infectious Mononucleosis. Ph.D. Thesis. University of Utah.

106. Taylor, R.N., H. Ehrhard and S. Marcus. 1976. Evaluation of tests for infectious mononucleosis through proficiency testing. Health Lab. Sci. 13:34-44.

107. Taylor, R.N., K.M. Fulford and A.Y. Huong. 1978. Comparison of kinetic and end-point diffusion methods for quantitation of immunoglobulins. J. Clin. Micro. 8:23-27.

108. Taylor, R.N., K.M. Fulford and A.Y. Huong. 1977. Results of a nationwide proficiency test for carcinoembryonic antigen. J. Clin. Micro. 5:433-438.

109. Taylor, R.N., K.M. Fulford and W.L. Jones. 1977. Reduction of variation in results of rheumatoid factor tests by use of a serum reference preparation. J. Clin. Micro. 5:42-45.

110. Taylor, R.N., K.M. Fulford, V.A. Przybyszewski and V. Pope. 1977. Center for Disease Control Diagnostic Immunology Proficiency Testing program results for 1976. J. Clin. Micro. 6:224-232.

111. Taylor, R.N., K.M. Fulford, V.A. Przybyszewski and V. Pope. 1978. Center for Disease Control Immunology Proficiency testing program results for 1977. J. Clin. Micro. 8:388-395.

112. Taylor, R.N., A.Y. Huong and K.M. Fulford. 1975. Measurement of Variation in Serologic Tests. Center for Disease Control, Atlanta, GA.

113. Todd, E.W. 1932. Antigenic streptococcal hemolysin J. Exptl. Med. 55:267-280.

114. Tonks, D.B. 1972. Quality Control in Clinical Laboratories. Warner-Chilcott Laboratories Co., Ltd., Scarborough, Ontario.

115. U.S. Department of Health, Education, and Welfare. (undated). The Morphologic and Serologic Diagnosis of Infectious Mononucleosis. National Communicable Disease Center, Atlanta, GA.

116. U.S. Department of Health, Education and Welfare. 1956-1957. Serology Evaluation and Research Assembly (SERA) Study. National Communicable Disease Center, Atlanta, GA.

117. U.S. Public Health Service. 1965. Standardized Diagnostic Complement Fixation Method and Adaptation to Micro Test. Public Health Monograph No. 74 (PHS Pub. No. 1228). U.S. Government Printing Office. Washington, D.C

118. Van Oss, C.J. 1973. Precipitation and agglutination p. 30-55. *In* Rose, N.R., F. Milgrom and C.J. van Oss (eds.). Principles of Immunology. Macmillan Publishing Co., New York, NY.

119. Van Oss, C.J. and P.M. Bronson. 1969. Immunorheophoresis. Immunochemistry 6:775-778.

120. Walker, W.H. 1977. An approach to immunoassay. Clin. Chem. 23:384.

121. Whitehead, T.P. 1977. Quality Control in Clinical Chemistry. John Wiley and Sons, New York, NY.

122. Wisdom, G.B. 1976. Enzyme-immunoassay. Clin. Chem. 22:1243-1255.

123. Wood, R.M. 1976. Tests for syphilis. p. 341-351. *In* Rose, N.R. and H. Friedman (eds.). Manual of Clinical Immunology. American Society for Microbiology, Washington, D.C.

124. World Health Organization. 1964. Joint FAO/WHO expert committee on brucellosis. Fourth Report. W.H.O. Tech. Report Series #289.

125. World Health Organization. 1970. W.H.O. expert committee on biological standardization. Twenty-second Report. W.H.O. Tech. Report Series #444.

126. World Health Organization. 1971. Joint FAO/WHO expert committee on brucellosis. Fifth Report. W.H.O. Tech. Report Series #464.

127. World Health Organization. 1973. W.H.O. expert committee on biological standardization. Twenty-fifth Report. W.H.O. Tech. Report Series #530.

128. World Health Organization. 1975. W.H.O. expert committee on biological standardization. Twenty-sixth Report. W.H.O. Tech. Report Series #565.

129. Wu, G.T., S.L. Twomey and R.E. Thiers. 1975. Statistical evaluation of method-comparison data. Clin. Chem. 21:315-320.

130. Yalow, R.S. and S.A. Berson. 1971. Introduction and general considerations p. 1. *In* Odell, W.D., W.H. Daughaday (eds.). Principles of Competitive Protein-Binding Assays. J.B. Lippincott Co.

131. Young, D.M. 1964. Experience with and thoughts on quality control. Am. J. Med. Technol. 30:99-122.

132. Zettner, A. 1973. Principles of competitive binding assays (saturation analyses). I. Equilibrium techniques. Clin. Chem. 19:699-705.

133. Zettner, A. and P.F. Duly. 1974. Principles of competitive binding assays (saturation analyses). II. Sequential saturation. Clin. Chem. 20:5-14.

GLOSSARY OF TERMS

Accuracy:

The relative exactness with which a sample estimate represents the true value. Exhibiting small bias. Sometimes used in discussion of multiple testing to refer to the combination of precision and lack of bias.

Bias Error:

A constant and false departure of a quantity from the actual value caused by error in test equipment or conditions.

Coefficient of Variation, Arithmetic (Relative Standard Deviation):

The standard deviation expressed as a percentage of the mean

$$(CV = \frac{S}{\bar{x}} \times 100)$$

Coefficient of Variation, Geometric:

(Not useful. Since the standard deviation limits are already a function of the mean, variation can be compared by direct examination of the standard deviation.)

Confidence Limits:

A set of mathematically established limits between which a specified percentage of the observations fall. For example, approximately 95% of the values from a Gaussian distribution will fall within the 2 S range (95% confidence limits).

Control Serum:

Serum with known concentration used to measure the accuracy and precision and sensitivity and specificity of a procedure. They are used to determine, verify and document performance.

Control Limits:

The limits for control results beyond which the test is considered invalid. They are determined by calculating the standard deviation and choosing the number of deviations which will produce the confidence limits desired.

Degrees of Freedom:

The number of independent observations in the sample (N), i.e. the sample size, minus the number of population parameters (K) which must be estimated from the sample observations. d.f. = N – K. Refer to the test procedure being used to determine how many estimates are made and consequently how many degrees of freedom remain.

Dilution:

The ratio of the volume of serum or other medium to the total volume produced by combining it with diluent. The dilution 1:10 denotes one part serum plus nine parts diluent for a total volume of 10 parts. A dilution of 1:1 indicates no dilution. The same notation is used in areas other than serology to indicate volume of material to volume of diluent. So the two methods must be carefully differentiated.

Efficiency:

The ability of a test to give a positive result on positives and negative results on negatives.

$$\text{Efficiency} = \frac{\text{True Positive} + \text{True Negative}}{\text{True Positive} + \text{False Positive} + \text{True Negative} + \text{False Negative}} \times 100.$$

Error:

The difference between an observed or measured value and the best obtainable estimates of its true value.

Experimental Error:

The deviation from the expected value attributed to the limitations of the procedure.

Frequency Distribution:

A tabular means of conveying large volumes of data by grouping the data into a number of classes with the frequency of results specified for each class.

Mean, Arithmetic:

The average of all observations. The sum of all observations divided by the number of observations ($\bar{x} = \frac{\Sigma x}{n}$). Because it is a theoretical value, the mean may not be encountered naturally.

Mean, Geometric:

The average of the logarithm of all observations, i.e., the arithmetic mean of log-transformed data, or, the nth root of the product of all observations $x_G = \sqrt[n]{x_1 \cdot x_2 \cdot x_3 \cdot \cdot \cdot x_n}$. A zero value will result in a zero mean; therefore, special consideration must be given to geometric data which contain zero values.

Median:

The middle value in an ordered array of values or the value which has an equal number of values above and below it, or the mean of the two middle values if there is no middle value.

BEST DOCUMENT AVAILABLE

Mode:
The value occurring with greatest frequency. The most common value.

Normal Distribution:
The theoretical, Gaussian, bell-shaped frequency distribution.

One-tailed (one-sided) test:
A test of the hypothesis that one process is better than another as opposed to testing whether one process is better or worse the another (which would be a two-tailed test). For example, the hypothesis "Test A results are not significantly higher than test B results" would call for a one-tailed test.

Outlier:
A result lying outside the general distribution of data. Can be excluded from further calculation on the basis of apparent invalidity. Also referred to as "obviously deviant" and is just that; a result that does not belong with the other results.

Parameter:
A quantity that describes a statistical sample or population. The normal distribution is completely described by the parameters: Mean (μ or $\overline{x}$) and standard deviation (σ or S).

Precision:
The closeness of the relationship between replicate determinations with no regard for bias error. Freedom from variation. Reproducibility. An inverse function of the variance; the greater the precision the smaller the variance.

Predictive Value:
Percentage of the positive results that are true positive, or percentage of the negative results that are true negative

$$\text{Predictive value of a positive} = \frac{\text{True Positive}}{\text{True Positive + False Positive}} \times 100.$$

$$\text{Predictive value of a negative} = \frac{\text{True Negative}}{\text{True Negative + False Negative}} \times 100.$$

Population:
A group of individual persons, objects, or items from which samples are taken for statistical measurement.

Probability:
Quantitative measure of chance. The ratio of the number of outcomes that produce a given event to the total number of possible outcomes.

Quality Control:
The system of activities intended to provide a quality product. A program in which known materials are used to confirm accuracy and precision,

sensitivity and specificity, detect errors, and thus check the reliability of results.

Random Error:

Error based on the inability to duplicate exact conditions. Tends to be normally distributed.

Randomization:

Any selection or assignment procedure designed to eliminate extraneous factors which might conceivably bias the results. Controls should be randomized within a testing sequence.

Range:

The difference between the maximum and minimum of a set of values.

Reference Material:

A material certified by a recognized standardization organization to be used as a basis for comparison.

Reliability:

The degree to which an analytic procedure maintains its accuracy, precision, specificity, and sensitivity over an extended period of time.

Repeatability:

The precision of replicate results of a single specimen tested at one time by the same person (within-run precision).

Reproducibility:

The precision of results when replicate results of a single specimen are tested by different persons at different times (among-run precision).

Sample:

A subset or group of objects or things selected from a larger set or population. Usually information obtained from the samples is used to make inferences about the population.

Sensitivity:

(1) The incidence of true positive results obtained when a test is used for patients known to have the disease or condition.

$$\text{Sensitivity} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}} \times 100.$$

(2) Ability to detect small quantities of substance being measured, or to detect small differences in concentration in a series of specimens (the slope of the response curve).

Shift:

A movement of values in either direction from the mean.

Significance Level:

A measure of the probability of an observed difference between an expected and observed value under the conditions of a hypothesized situation.

Specificity:

A measure of the ability of a test to indicate the absence of a component in a specimen when it is truly negative for that component, or a measure of the ability of a test to measure accurately one component in a specimen without interference by other components.

(1) A measure of the ability of a test to give a negative result in the absence of a disease.

(2) $\text{Specificity} = \dfrac{\text{True Negative}}{\text{True Negative + False Positive}} \times 100.$

Standard (calibration standard or calibrator):

A substance of known concentration (usually determined by comparison to a reference material) used as a basis for determining the concentration of the substance in unknown specimens.

Standard Curve:

A curve obtained by testing various concentrations of a standard material (calibrator) which is used to determine the concentration in unknown materials.

Standard Deviation, Arithmetic:

The square root of the average of squared deviations from the arithmetic mean $S = \sqrt{\dfrac{\Sigma(x-\bar{x})^2}{n-1}}$. The standard deviation provides a measure of the deviation of values about the mean in the units of the observations. When $n \geq 30$ either n or $n-1$ may be used in the denominator since they are practically equal. The percentages of observations from a Gaussian distribution included within standard deviation limits are as follows: 1 S = 68.27%, 2 S = 95.45%, and 3 S = 99.73%.

Standard Deviation, Geometric:

The antilog of the standard deviation of log transformed data.

$$S_G = \frac{\Sigma(\log x - \log \bar{x})^2}{n-1}$$

The percentages of observations within standard deviation limits are the same as those for the normal distribution, but the limits are not symmetrical about the geometric mean.

Titer, Measured:

The *reciprocal* of the highest dilution of a constituent which leads to a positive reaction. For example, if a serum produces a positive reaction in the tube which contains 1:80 dilution but not in the next higher dilution, then it has a titer of the *reciprocal* of 1:80, or 80. The dilution is 1:80; the titer is 80. A titer of 80 means that the serum is positive in the 1:80 dilution.

Titer, True:

The actual titer of the constituent being measured without regard to the dilution ratio of the system of measurement. Frequently the true titer falls between possible values in the dilution system, it will therefore, be higher than the measured titer.

Trend:

The tendency of values to move successively further from the mean.

Two-tailed (two-sided) test:

A test of the hypothesis that a process is either better or worse than another as opposed to testing whether one process is better than another (which would be a one-tailed test). For example, the hypothesis "Test A results are not significantly different (neither higher nor lower) from Test B results" would call for a two-tailed test.

Variance:

A measure of the deviation of values about the mean. The standard deviation squared (S^2). The total variance in a procedure can be separated into components of variance. Variances are additive, but standard deviations are not. Variance is an inverse function of the precision; the greater the precision, the smaller the variance. Variance for geometric data should be calculated from log-transformed data, and all data to be compared must be log-transformed to the same base.

APPENDICES

BLANK PAGE

Appendix A

Cumulative Normal Frequency Distribution
(Area under the standard normal curve from 0 to Z)

Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	.0398	.0438	.0478	.0517	.0557	.0596	.0636	.0675	.0714	.0753
0.2	.0793	.0832	.0871	.0910	.0948	.0987	.1026	.1064	.1103	.1141
0.3	.1179	.1217	.1255	.1293	.1331	.1368	.1406	.1443	.1480	.1517
0.4	.1554	.1591	.1628	.1664	.1700	.1736	.1772	.1808	.1844	.1879
0.5	.1915	.1950	.1985	.2019	.2054	.2088	.2123	.2157	.2190	.2224
0.6	.2257	.2291	.2324	.2357	.2389	.2422	.2454	.2486	.2517	.2549
0.7	.2580	.2611	.2642	.2673	.2704	.2734	.2764	.2794	.2823	.2852
0.8	.2881	.2910	.2939	.2967	.2995	.3023	.3051	.3078	.3106	.3133
0.9	.3159	.3186	.3212	.3238	.3264	.3289	.3315	.3340	.3365	.3389
1.0	.3413	.3438	.3461	.3485	.3508	.3531	.3554	.3577	.3599	.3621
1.1	.3643	.3665	.3686	.3708	.3729	.3749	.3770	.3790	.3810	.3830
1.2	.3849	.3869	.3888	.3907	.3925	.3944	.3962	.3980	.3997	.4015
1.3	.4032	.4049	.4066	.4082	.4099	.4115	.4131	.4147	.4162	.4177
1.4	.4192	.4207	.4222	.4236	.4251	.4265	.4279	.4292	.4306	.4319
1.5	.4332	.4345	.4357	.4370	.4382	.4394	.4406	.4418	.4429	.4441
1.6	.4452	.4463	.4474	.4484	.4495	.4505	.4515	.4525	.4535	.4545
1.7	.4554	.4564	.4573	.4582	.4591	.4599	.4608	.4616	.4625	.4633
1.8	.4641	.4649	.4656	.4664	.4671	.4678	.4686	.4693	.4699	.4706
1.9	.4713	.4719	.4726	.4732	.4738	.4744	.4750	.4756	.4761	.4767
2.0	.4772	.4778	.4783	.4788	.4793	.4798	.4803	.4808	.4812	.4817
2.1	.4821	.4826	.4830	.4834	.4838	.4842	.4846	.4850	.4854	.4857
2.2	.4861	.4864	.4868	.4871	.4875	.4878	.4881	.4884	.4887	.4890
2.3	.4893	.4896	.4898	.4901	.4904	.4906	.4909	.4911	.4913	.4916
2.4	.4918	.4920	.4922	.4925	.4927	.4929	.4931	.4932	.4934	.4936
2.5	.4938	.4940	.4941	.4943	.4945	.4946	.4948	.4949	.4951	.4952
2.6	.4953	.4955	.4956	.4957	.4959	.4960	.4961	.4962	.4963	.4964
2.7	.4965	.4966	.4967	.4968	.4969	.4970	.4971	.4972	.4973	.4974
2.8	.4974	.4975	.4976	.4977	.4977	.4978	.4979	.4979	.4980	.4981
2.9	.4981	.4982	.4982	.4983	.4984	.4984	.4985	.4985	.4986	.4986
3.0	.4987	.4987	.4987	.4988	.4988	.4989	.4989	.4989	.4990	.4990
3.1	.4990	.4991	.4991	.4991	.4992	.4992	.4992	.4992	.4993	.4993
3.2	.4993	.4993	.4994	.4994	.4994	.4994	.4994	.4995	.4995	.4995
3.3	.4995	.4995	.4995	.4996	.4996	.4996	.4996	.4996	.4996	.4997
3.4	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4998
3.6	.4998	.4998	.4999	.4999	.4999	.4999	.4999	.4999	.4999	.4999
3.9	.5000									

Reprinted from *Statistical Methods* by G. W. Snedecor and W. G. Cochran, by permission of Iowa State University Press, Ames, Iowa.

Appendix B

Cumulative Distribution of Chi-Square

Degrees of Freedom	Probability of a Greater Value					
	0.250	0.100	0.050	0.025	0.010	0.005
1	1.32	2.71	3.84	5.02	6.63	7.88
2	2.77	4.61	5.99	7.38	9.21	10.60
3	4.11	6.25	7.81	9.35	11.34	12.84
4	5.39	7.78	9.49	11.14	13.28	14.86
5	6.63	9.24	11.07	12.83	15.09	16.75
6	7.84	10.64	12.59	14.45	16.81	18.55
7	9.04	12.02	14.07	16.01	18.48	20.28
8	10.22	13.36	15.51	17.53	20.09	21.96
9	11.39	14.68	16.92	19.02	21.67	23.59
10	12.55	15.99	18.31	20.48	23.21	25.19
11	13.70	17.28	19.68	21.92	24.72	26.76
12	14.85	18.55	21.03	23.34	26.22	28.30
13	15.98	19.81	22.36	24.74	27.69	29.82
14	17.12	21.06	23.68	26.12	29.14	31.32
15	18.25	22.31	25.00	27.49	30.58	32.80
16	19.37	23.54	26.30	28.85	32.00	34.27
17	20.49	24.77	27.59	30.19	33.41	35.72
18	21.60	25.99	28.87	31.53	34.81	37.16
19	22.72	27.20	30.14	32.85	36.19	38.58
20	23.83	28.41	31.41	34.17	37.57	40.00
21	24.93	29.62	32.67	35.48	38.92	41.40
22	26.04	30.81	33.92	36.78	40.29	42.80
23	27.14	32.01	35.17	38.08	41.64	44.18
24	28.24	33.20	36.42	39.36	42.98	45.56
25	29.34	34.38	37.65	40.65	44.31	46.93
26	30.43	35.56	38.89	41.92	45.64	48.29
27	31.53	36.74	40.11	43.19	46.96	49.64
28	32.62	37.92	41.34	44.46	48.28	50.99
29	33.71	39.09	42.56	45.72	49.59	52.34
30	34.80	40.26	43.77	46.98	50.89	53.67
40	45.62	51.80	55.76	59.34	63.69	66.77
50	56.33	63.17	67.50	71.42	76.15	79.49
60	66.98	74.40	79.08	83.30	88.38	91.95
70	77.58	85.53	90.53	95.02	100.42	104.22
80	88.13	96.58	101.88	106.63	112.33	116.32
90	98.64	107.56	113.14	118.14	124.12	128.30
100	109.14	118.50	124.34	129.56	135.81	140.17

Reprinted from *Statistical Methods* by G. W. Snedecor and W. G. Cochran, by permission of Iowa State University Press, Ames, Iowa.

Appendix C

Table of Critical Values of *D* In The Kolmogorov-Smirnov Goodness of Fit Test

Sample Size (n)	Level of Significance for D=Maximum $\lvert F(x)-S_n(x)\rvert$				
	.20	.15	.10	.05	.01
1	.900	.925	.950	.975	.995
2	.684	.726	.776	.842	.929
3	.565	.597	.642	.708	.828
4	.494	.525	.564	.624	.733
5	.446	.474	.510	.565	.669
6	.410	.436	.470	.521	.618
7	.381	.405	.438	.486	.577
8	.358	.381	.411	.457	.543
9	.339	.360	.388	.432	.514
10	.322	.342	.368	.410	.490
11	.307	.326	.352	.391	.468
12	.295	.313	.338	.375	.450
13	.284	.302	.325	.361	.433
14	.274	.292	.314	.349	.418
15	.266	.283	.304	.338	.404
16	.258	.274	.295	.328	.392
17	.250	.266	.286	.318	.381
18	.244	.259	.278	.309	.371
19	.237	.252	.272	.301	.363
20	.231	.246	.264	.294	.356
25	.21	.22	.24	.27	.32
30	.19	.20	.22	.24	.29
35	.18	.19	.21	.23	.27
Over 35	$\frac{1.07}{\sqrt{n}}$	$\frac{1.14}{\sqrt{n}}$	$\frac{1.22}{\sqrt{n}}$	$\frac{1.36}{\sqrt{n}}$	$\frac{1.63}{\sqrt{n}}$

Adapted from F. J. Massey, Jr., "The Kolmogorov-Smirnov test for goodness of fit," *Jour. Amer. Stat. Assn.*, Vol. 46, 1951, pp. 68-78. By permission of the author and publisher.

The values of D given in the table are critical values associated with selected values of n. Any value of D which is greater than or equal to the tabulated value is significant at the indicated level of significance.

Appendix C *(Continued)*

Criteria for Testing for Extreme Mean

Statistic	Number of obs., k	Critical values						
		$\alpha =$ .30	$\alpha =$ .20	$\alpha =$ .10	$\alpha =$ .05	$\alpha =$ .02	$\alpha =$ .01	$\alpha =$ .005
$r_{10} = \frac{X_2 - X_1}{X_k - X_1}$	3	.684	.781	.886	.941	.976	.988	.994
	4	.471	.560	.679	.765	.846	.889	.926
	5	.373	.451	.557	.642	.729	.780	.821
	6	.318	.386	.482	.560	.644	.698	.740
	7	.281	.344	.434	.507	.586	.637	.680
$r_{11} = \frac{X_2 - X_1}{X_{k-1} - X_1}$	8	.318	.385	.479	.554	.631	.683	.725
	9	.288	.352	.441	.512	.587	.635	.677
	10	.265	.325	.409	.477	.551	.597	.639
$r_{21} = \frac{X_3 - X_1}{X_{k-1} - X_1}$	11	.391	.442	.517	.576	.638	.679	.713
	12	.370	.419	.490	.546	.605	.642	.675
	13	.351	.399	.467	.521	.578	.615	.649
$r_{22} = \frac{X_3 - X_1}{X_{k-2} - X_1}$	14	.370	.421	.492	.546	.602	.641	.674
	15	.353	.402	.472	.525	.579	.616	.647
	16	.338	.386	.454	.507	.559	.595	.624
	17	.325	.373	.438	.490	.542	.577	.605
	18	.314	.361	.424	.475	.527	.561	.589
	19	.304	.350	.412	.462	.514	.547	.575
	20	.295	.340	.401	.450	.502	.535	.562
	21	.287	.331	.391	.440	.491	.524	.551
	22	.280	.323	.382	.430	.481	.514	.541
	23	.274	.316	.374	.421	.472	.505	.532
	24	.268	.310	.367	.413	.464	.497	.524
	25	.262	.304	.360	.406	.457	.489	.516

Adapted from F. J. Massey, Jr., "The Kolmogorov-Smirnov test for goodness of fit," *Jour. Amer. Stat. Assn.*, Vol. 46, 1951, pp. 68-78. By permission of the author and publisher.

APPENDIX D

Distribution of t

	Probability of Larger Value Sign Ignored						
One Tailed Test	0.1000	0.0500	0.0250	0.0125	0.0050	0.0025	0.0005
Two Tailed Test	0.200	0.100	0.050	0.025	0.010	0.005	0.001
Degrees of Freedom							
1	3.078	6.314	12.706	25.452	63.657		
2	1.886	2.920	4.303	6.205	9.925	14.089	31.598
3	1.638	2.353	3.182	4.176	5.841	7.453	12.941
4	1.533	2.132	2.776	3.495	4.604	5.598	8.610
5	1.476	2.015	2.571	3.163	4.032	4.773	6.859
6	1.440	1.943	2.447	2.969	3.707	4.317	5.959
7	1.415	1.895	2.365	2.841	3.499	4.029	5.405
8	1.397	1.860	2.306	2.752	3.355	3.832	5.041
9	1.383	1.833	2.262	2.685	3.250	3.690	4.781
10	1.372	1.812	2.228	2.634	3.169	3.581	4.587
11	1.363	1.796	2.201	2.593	3.106	3.497	4.437
12	1.356	1.782	2.179	2.560	3.055	3.428	4.318
13	1.350	1.771	2.160	2.533	3.012	3.372	4.221
14	1.345	1.761	2.145	2.510	2.977	3.326	4.140
15	1.341	1.753	2.131	2.490	2.947	3.286	4.073
16	1.337	1.746	2.120	2.473	2.921	3.252	4.015
17	1.333	1.740	2.110	2.458	2.898	3.222	3.965
18	1.330	1.734	2.101	2.445	2.878	3.197	3.922
19	1.328	1.729	2.093	2.433	2.861	3.174	3.883
20	1.325	1.725	2.086	2.423	2.845	3.153	3.850
21	1.323	1.721	2.080	2.414	2.831	3.135	3.819
22	1.321	1.717	2.074	2.406	2.819	3.119	3.792
23	1.319	1.714	2.069	2.398	2.807	3.104	3.767
24	1.318	1.711	2.064	2.391	2.797	3.090	3.745
25	1.316	1.708	2.060	2.385	2.787	3.078	3.725
26	1.315	1.706	2.056	2.379	2.779	3.067	3.707
27	1.314	1.703	2.052	2.373	2.771	3.056	3.690
28	1.313	1.701	2.048	2.368	2.763	3.047	3.674
29	1.311	1.699	2.045	2.364	2.756	3.038	3.659
30	1.310	1.697	2.042	2.360	2.750	3.030	3.646
35	1.306	1.690	2.030	2.342	2.724	2.996	3.591
40	1.303	1.684	2.021	2.329	2.704	2.971	3.551
45	1.301	1.680	2.014	2.319	2.690	2.952	3.520
50	1.299	1.676	2.008	2.310	2.678	2.937	3.496
55	1.297	1.673	2.004	2.304	2.669	2.925	3.476
60	1.296	1.671	2.000	2.299	2.660	2.915	3.460
70	1.294	1.667	1.994	2.290	2.648	2.899	3.435
80	1.293	1.665	1.989	2.284	2.638	2.887	3.416
90	1.291	1.662	1.986	2.279	2.631	2.878	3.402
100	1.290	1.661	1.982	2.276	2.625	2.871	3.390
120	1.289	1.658	1.980	2.270	2.617	2.860	3.373
∞	1.2816	1.6448	1.9600	2.2414	2.5758	2.8070	3.2905

Adapted from *Statistical Methods* by G. W. Snedecor and W. G. Cochran, by permission of Iowa State University Press, Ames, IA and from *Statistical Methods for Research Workers* (14th Ed.) by R. A. Fisher, by permission of MacMillan Publishing Co., New York, NY.

APPENDIX E

Distribution of F

5% (Upper Line) and 1% (Lower Line) Points for the Distribution of *F*

f_2	f_1 Degrees of Freedom (for greater mean squares)																								f_2
	1	2	3	4	5	6	7	8	9	10	11	12	14	16	20	24	30	40	50	75	100	200	500	∞	
1	161	200	216	225	230	234	237	239	241	242	243	244	245	246	248	249	250	251	252	253	253	254	254	254	1
	4,052	4,999	5,403	5,625	5,764	5,859	5,928	5,981	6,022	6,056	6,082	6,106	6,142	6,169	6,208	6,234	6,261	6,286	6,302	6,323	6,334	6,352	6,361	6,366	
2	18.51	19.00	19.16	19.25	19.30	19.33	19.36	19.37	19.38	19.39	19.40	19.41	19.42	19.43	19.44	19.45	19.46	19.47	19.47	19.48	19.49	19.49	19.50	19.50	2
	98.49	99.00	99.17	99.25	99.30	99.33	99.36	99.37	99.39	99.40	99.41	99.42	99.43	99.44	99.45	99.46	99.47	99.48	99.48	99.49	99.49	99.49	99.50	99.50	
3	10.13	9.55	9.28	9.12	9.01	8.94	8.88	8.84	8.81	8.78	8.76	8.74	8.71	8.69	8.66	8.64	8.62	8.60	8.58	8.57	8.56	8.54	8.54	8.53	3
	34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.49	27.34	27.23	27.13	27.05	26.92	26.83	26.69	26.60	26.50	26.41	26.35	26.27	26.23	26.13	26.14	26.12	
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.93	5.91	5.87	5.84	5.80	5.77	5.74	5.71	5.70	5.68	5.66	5.65	5.64	5.63	4
	21.20	18.00	16.69	15.98	15.52	15.21	14.98	14.80	14.66	14.54	14.45	14.37	14.24	14.15	14.02	13.93	13.83	13.74	13.69	13.61	13.57	13.52	13.48	13.46	
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.78	4.74	4.70	4.68	4.64	4.60	4.56	4.53	4.50	4.46	4.44	4.42	4.40	4.38	4.37	4.36	5
	16.26	13.27	12.06	11.39	10.97	10.67	10.45	10.29	10.15	10.05	9.96	9.89	9.77	9.68	9.55	9.47	9.38	9.29	9.24	9.17	9.13	9.07	9.04	9.02	
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.03	4.00	3.96	3.92	3.87	3.84	3.81	3.77	3.75	3.72	3.71	3.69	3.68	3.67	6
	13.74	10.92	9.78	9.15	8.75	8.47	8.26	8.10	7.98	7.87	7.79	7.72	7.60	7.52	7.39	7.31	7.23	7.14	7.09	7.02	6.99	6.94	6.90	6.88	
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.63	3.60	3.57	3.52	3.49	3.44	3.41	3.38	3.34	3.32	3.29	3.28	3.25	3.24	3.23	7
	12.25	9.55	8.45	7.85	7.46	7.19	7.00	6.84	6.71	6.62	6.65	6.47	6.35	6.27	6.15	6.07	5.98	5.90	5.85	5.78	5.75	5.70	5.67	5.65	
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.34	3.31	3.28	3.23	3.20	3.15	3.12	3.08	3.05	3.03	3.00	2.98	2.96	2.94	2.93	8
	11.26	8.65	7.59	7.01	6.63	6.37	6.19	6.03	5.91	5.82	5.74	5.67	5.56	5.48	5.36	5.28	5.20	5.11	5.06	5.00	4.96	4.91	4.88	4.86	
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.13	3.10	3.07	3.02	2.98	2.93	2.90	2.86	2.82	2.80	2.77	2.76	2.73	2.72	2.71	9
	10.56	8.02	6.99	6.42	6.06	5.80	5.62	5.47	5.35	5.26	5.18	5.11	5.00	4.92	4.80	4.73	4.64	4.56	4.51	4.45	4.41	4.36	4.33	4.31	
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.97	2.94	2.91	2.86	2.82	2.77	2.74	2.70	2.67	2.64	2.61	2.59	2.56	2.55	2.54	10
	10.04	7.56	6.55	5.99	5.64	5.39	5.21	5.06	4.95	4.85	4.78	4.71	4.60	4.52	4.41	4.33	4.25	4.17	4.12	4.05	4.01	3.96	3.93	3.91	
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.86	2.82	2.79	2.74	2.70	2.65	2.61	2.57	2.53	2.50	2.47	2.45	2.42	2.41	2.40	11
	9.65	7.20	6.22	5.67	5.32	5.07	4.88	4.74	4.63	4.54	4.46	4.40	4.29	4.21	4.10	4.02	3.94	3.86	3.80	3.74	3.70	3.66	3.62	3.60	
12	4.75	3.88	3.49	3.26	3.11	3.00	2.92	2.85	2.80	2.76	2.72	2.69	2.64	2.60	2.54	2.50	2.46	2.42	2.40	2.36	2.35	2.32	2.31	2.30	12
	9.33	6.93	5.95	5.41	5.06	4.82	4.65	4.50	4.39	4.30	4.22	4.16	4.05	3.98	3.86	3.78	3.70	3.61	3.56	3.49	3.46	3.41	3.38	3.36	
13	4.67	3.80	3.41	3.18	3.02	2.92	2.84	2.77	2.72	2.67	2.63	2.60	2.55	2.51	2.46	2.42	2.38	2.34	2.32	2.28	2.26	2.24	2.22	2.21	13
	9.07	6.70	5.74	5.20	4.86	4.62	4.44	4.30	4.19	4.10	4.02	3.96	3.85	3.78	3.67	3.59	3.51	3.42	3.37	3.30	3.27	3.21	3.18	3.16	

Distribution of F (continued)

f_2	f_1 Degrees of Freedom (greater mean square)																								f_2
	1	2	3	4	5	6	7	8	9	10	11	12	14	16	20	24	30	40	50	75	100	200	500	∞	
14	4.60	3.74	3.34	3.11	2.96	2.85	2.77	2.70	2.65	2.60	2.56	2.53	2.48	2.44	2.39	2.35	2.31	2.27	2.24	2.21	2.19	2.16	2.14	2.13	14
	8.86	6.51	5.56	5.03	4.69	4.46	4.28	4.14	4.03	3.94	3.86	3.80	3.70	3.62	3.51	3.43	3.34	3.26	3.21	3.14	3.11	3.06	3.02	3.00	
15	4.54	3.68	3.29	3.06	2.90	2.79	2.70	2.64	2.59	2.55	2.51	2.48	2.43	2.39	2.33	2.29	2.25	2.21	2.18	2.15	2.12	2.10	2.08	2.07	15
	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89	3.80	3.73	3.67	3.56	3.48	3.36	3.29	3.20	3.12	3.07	3.00	2.97	2.92	2.89	2.87	
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.45	2.42	2.37	2.33	2.28	2.24	2.20	2.16	2.13	2.09	2.07	2.04	2.02	2.01	16
	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78	3.69	3.61	3.55	3.45	3.37	3.25	3.18	3.10	3.01	2.96	2.98	2.86	2.80	2.77	2.75	
17	4.45	3.59	3.20	2.96	2.81	2.70	2.62	2.55	2.50	2.45	2.41	2.38	2.33	2.29	2.23	2.19	2.15	2.11	2.08	2.04	2.02	1.99	1.97	1.96	17
	8.40	6.11	5.18	4.67	4.34	4.10	3.93	3.79	3.68	3.59	3.52	3.45	3.35	3.27	3.16	3.08	3.00	2.92	2.86	2.79	2.76	2.70	2.67	2.65	
18	4.41	3.55	3.16	2.93	2.77	3.66	2.58	2.51	2.46	2.41	2.37	2.34	2.29	2.25	2.19	2.15	2.11	2.07	2.04	2.00	1.98	1.95	1.93	1.92	18
	8.28	6.01	5.09	4.58	4.25	4.01	3.85	3.71	3.60	3.51	3.44	3.37	3.27	3.19	3.07	3.00	2.91	2.83	2.78	2.71	2.68	2.62	2.59	2.57	
19	4.38	3.52	3.13	2.90	2.74	2.63	2.55	2.48	2.43	2.38	2.34	2.31	2.26	2.21	2.15	2.11	2.07	2.02	2.00	1.96	1.94	1.91	1.90	1.88	19
	8.18	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52	3.43	3.36	3.30	3.19	3.12	3.00	2.92	2.84	2.76	2.70	2.63	2.60	2.54	2.51	2.49	
20	4.35	3.49	3.10	2.87	2.71	2.60	2.52	2.45	2.40	2.35	2.31	2.28	2.23	2.18	2.12	2.08	2.04	1.99	1.96	1.92	1.90	1.87	1.85	1.84	20
	8.10	5.85	4.94	4.43	4.10	3.87	3.71	3.56	3.45	3.37	3.30	3.23	3.13	3.05	2.94	2.86	2.77	2.69	2.63	2.56	2.53	2.47	2.44	2.42	
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32	2.28	2.25	2.20	2.15	2.09	2.05	2.00	1.96	1.93	1.89	1.87	1.84	1.82	1.81	21
	8.02	5.78	4.87	4.37	4.04	3.81	3.65	3.51	3.40	3.31	3.24	3.17	3.07	2.99	2.88	2.80	2.72	2.63	2.58	2.51	2.47	2.42	2.38	2.36	
22	4.30	3.44	3.05	2.82	2.66	2.55	2.47	2.40	2.35	2.30	2.26	2.23	2.18	2.13	2.07	2.03	1.98	1.93	1.91	1.87	1.84	1.81	1.80	1.78	22
	7.94	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35	3.26	3.18	3.12	3.02	2.94	2.83	2.75	2.67	2.58	2.53	2.46	2.42	2.37	2.33	2.31	
23	4.28	3.42	3.03	2.80	2.64	2.53	2.45	2.38	2.32	2.28	2.24	2.20	2.14	2.10	2.04	2.00	1.96	1.91	1.88	1.84	1.82	1.79	1.77	1.76	23
	7.88	5.66	4.76	4.26	3.94	3.71	3.54	3.41	3.30	3.21	3.14	3.07	2.97	2.89	2.78	2.70	2.62	2.53	2.48	2.41	2.37	2.32	2.28	2.26	
24	4.26	3.40	3.01	2.78	2.62	2.51	2.43	2.36	2.30	2.26	2.22	2.18	2.13	2.09	2.02	1.98	1.94	1.89	1.86	1.82	1.80	1.76	1.74	1.73	24
	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.25	3.17	3.09	3.03	2.93	2.85	2.74	2.66	2.58	2.49	2.44	2.36	2.33	2.27	2.23	2.21	
25	4.24	3.38	2.99	2.76	2.60	2.49	2.41	2.34	2.28	2.24	2.20	2.16	2.11	2.06	2.00	1.96	1.92	1.87	1.84	1.80	1.77	1.74	1.72	1.71	25
	7.77	5.57	4.68	4.18	3.86	3.63	3.46	3.32	3.21	3.13	3.05	2.99	2.89	2.81	2.70	2.62	2.54	2.45	2.40	2.32	2.29	2.23	2.19	2.17	
26	4.22	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22	2.18	2.15	2.10	2.05	1.99	1.95	1.90	1.85	1.82	1.78	1.76	1.72	1.70	1.69	26
	7.72	5.53	4.64	4.14	3.82	3.59	3.42	3.29	3.17	3.09	3.02	2.96	2.86	2.77	2.66	2.58	2.50	2.41	2.36	2.28	2.25	2.19	2.15	2.13	

APPENDIX E (continued)

Distribution of F (continued)

f_2	f_1 Degrees of Freedom (for greater mean square)																								f_2
	1	2	3	4	5	6	7	8	9	10	11	12	14	16	20	24	30	40	50	75	100	200	500	∞	
27	4.21 7.68	3.35 5.49	2.96 4.60	2.73 4.11	2.57 3.79	2.46 3.56	2.37 3.39	2.30 3.26	2.25 3.14	2.20 3.06	2.16 2.98	2.13 2.93	2.08 2.83	2.03 2.74	1.97 2.63	1.93 2.55	1.88 2.47	1.84 2.38	1.80 2.33	1.76 2.25	1.74 2.21	1.71 2.16	1.68 2.12	1.67 2.10	27
28	4.20 7.64	3.34 5.45	2.95 4.57	2.71 4.07	2.56 3.76	2.44 3.53	2.36 3.36	2.29 3.23	2.24 3.11	2.19 3.03	2.15 2.95	2.12 2.90	2.06 2.80	2.02 2.71	1.96 2.60	1.91 2.52	1.87 2.44	1.81 2.35	1.78 2.30	1.75 2.22	1.72 2.18	1.69 2.13	1.67 2.09	1.65 2.06	28
29	4.18 7.60	3.33 5.42	2.93 4.54	2.70 4.04	2.54 3.73	2.43 3.50	2.35 3.33	2.28 3.20	2.22 3.08	2.18 3.00	2.14 2.92	2.10 2.87	2.05 2.77	2.00 2.68	1.94 2.57	1.90 2.49	1.85 2.41	1.80 2.32	1.77 2.27	1.73 2.19	1.71 2.15	1.68 2.10	1.65 2.06	1.64 2.03	29
30	4.17 7.56	3.32 5.39	2.92 4.51	2.69 4.02	2.53 3.70	2.42 3.47	2.34 3.30	2.27 3.17	2.21 3.06	2.16 2.98	2.12 2.90	2.09 2.84	2.04 2.74	1.99 2.66	1.93 2.55	1.89 2.47	1.84 2.38	1.79 2.29	1.76 2.24	1.72 2.16	1.69 2.13	1.66 2.07	1.64 2.03	1.62 2.01	30
32	4.15 7.50	3.30 5.34	2.90 4.46	2.67 3.97	2.51 3.66	2.40 3.42	2.32 3.25	2.25 3.12	2.19 3.01	2.14 2.94	2.10 2.86	2.07 2.80	2.02 2.70	1.97 2.62	1.91 2.51	1.86 2.42	1.82 2.34	1.76 2.25	1.74 2.20	1.69 2.12	1.67 2.08	1.64 2.02	1.61 1.98	1.59 1.96	32
34	4.13 7.44	3.28 5.29	2.88 4.42	2.65 3.93	2.49 3.61	2.38 3.38	2.30 3.21	2.23 3.08	2.17 2.97	2.12 2.89	2.08 2.82	2.05 2.76	2.00 2.66	1.95 2.58	1.89 2.47	1.84 2.38	1.80 2.30	1.74 2.21	1.71 2.15	1.67 2.08	1.64 2.04	1.61 1.98	1.59 1.94	1.57 1.91	34
36	4.11 7.39	3.26 5.25	2.86 4.38	2.63 3.89	2.48 3.58	2.36 3.35	2.28 3.18	2.21 3.04	2.15 2.94	2.10 2.86	2.06 2.78	2.03 2.72	1.98 2.62	1.93 2.54	1.87 2.43	1.82 2.35	1.78 2.26	1.72 2.17	1.69 2.12	1.65 2.04	1.62 2.00	1.59 1.94	1.56 1.90	1.55 1.87	36
38	4.10 7.35	3.25 5.21	2.85 4.34	2.62 3.86	2.46 3.54	2.35 3.32	2.26 3.15	2.19 3.02	2.14 2.91	2.09 2.82	2.05 2.75	2.02 2.69	1.96 2.59	1.92 2.51	1.85 2.40	1.80 2.32	1.76 2.22	1.71 2.14	1.67 2.08	1.63 2.00	1.60 1.97	1.57 1.90	1.54 1.86	1.53 1.84	38
40	4.08 7.31	3.23 5.18	2.84 4.31	2.61 3.83	2.45 3.51	2.34 3.29	2.25 3.12	2.18 2.99	2.12 2.88	2.07 2.80	2.04 2.73	2.00 2.66	1.95 2.56	1.90 2.49	1.84 2.37	1.79 2.29	1.74 2.20	1.69 2.11	1.66 2.05	1.61 1.97	1.59 1.94	1.55 1.88	1.53 1.84	1.51 1.81	40
42	4.07 7.27	3.22 5.15	2.83 4.29	2.59 3.80	2.44 3.49	2.32 3.26	2.24 3.10	2.17 2.96	2.11 2.86	2.06 2.77	2.02 2.70	1.99 2.64	1.94 2.54	1.89 2.46	1.82 2.35	1.78 2.26	1.73 2.17	1.68 2.08	1.64 2.02	1.60 1.94	1.57 1.91	1.54 1.85	1.51 1.80	1.49 1.78	42
44	4.06 7.24	3.21 5.12	2.82 4.26	2.58 3.78	2.43 3.46	2.31 3.24	2.23 3.07	2.16 2.94	2.10 2.84	2.05 2.75	2.01 2.68	1.98 2.62	1.92 2.52	1.88 2.44	1.81 2.32	1.76 2.24	1.72 2.15	1.66 2.06	1.63 2.00	1.58 1.92	1.56 1.88	1.52 1.82	1.50 1.78	1.48 1.75	44
46	4.05 7.21	3.20 5.10	2.81 4.24	2.57 3.76	2.42 3.44	2.30 3.22	2.22 3.05	2.14 2.92	2.09 2.82	2.04 2.73	2.00 2.66	1.97 2.60	1.91 2.50	1.87 2.42	1.80 2.30	1.75 2.22	1.71 2.13	1.65 2.04	1.62 1.98	1.57 1.90	1.54 1.86	1.51 1.80	1.48 1.76	1.46 1.72	46
48	4.04 7.19	3.19 5.08	2.80 4.22	2.56 3.74	2.41 3.42	2.30 3.20	2.21 3.04	2.14 2.90	2.08 2.80	2.03 2.71	1.99 2.64	1.96 2.58	1.90 2.48	1.86 2.40	1.79 2.28	1.74 2.20	1.70 2.11	1.64 2.02	1.61 1.96	1.56 1.88	1.53 1.84	1.50 1.78	1.47 1.73	1.45 1.70	48

APPENDIX E (continued)

Distribution of F (continued)

f_2	f_1 Degrees of Freedom (for greater mean square)																								f_2
	1	2	3	4	5	6	7	8	9	10	11	12	14	16	20	24	30	40	50	75	100	200	500	∞	
50	4.03	3.18	2.79	2.56	2.40	2.29	2.20	2.13	2.07	2.02	1.98	1.95	1.90	1.85	1.78	1.74	1.69	1.63	1.60	1.55	1.52	1.48	1.46	1.44	50
	7.17	5.06	4.20	3.72	3.41	3.18	3.02	2.88	2.78	2.70	2.62	2.56	2.46	2.39	2.26	2.18	2.10	2.00	1.94	1.86	1.82	1.76	1.71	1.68	
55	4.02	3.17	2.78	2.54	2.38	2.27	2.18	2.11	2.05	2.00	1.97	1.93	1.88	1.83	1.76	1.72	1.67	1.61	1.58	1.52	1.50	1.46	1.43	1.41	55
	7.12	5.01	4.16	3.68	3.37	3.15	2.98	2.85	2.75	2.66	2.59	2.53	2.43	2.35	2.23	2.15	2.06	1.96	1.90	1.82	1.78	1.71	1.66	1.64	
60	4.00	3.15	2.76	2.52	2.37	2.25	2.17	2.10	2.04	1.99	1.95	1.92	1.86	1.81	1.75	1.70	1.65	1.59	1.56	1.50	1.48	1.44	1.41	1.39	60
	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72	2.63	2.56	2.50	2.40	2.32	2.20	2.12	2.03	1.93	1.87	1.79	1.74	1.68	1.63	1.60	
65	3.99	3.14	2.75	2.51	2.36	2.24	2.15	2.08	2.02	1.98	1.94	1.90	1.85	1.80	1.73	1.68	1.63	1.57	1.54	1.49	1.46	1.42	1.39	1.37	65
	7.04	4.95	4.10	3.62	3.31	3.09	2.93	2.79	2.70	2.61	2.54	2.47	2.37	2.30	2.18	2.09	2.00	1.90	1.84	1.76	1.71	1.64	1.60	1.56	
70	3.98	3.13	2.74	2.50	2.35	2.23	2.14	2.07	2.01	1.97	1.93	1.89	1.84	1.79	1.72	1.67	1.62	1.56	1.53	1.47	1.45	1.40	1.37	1.35	70
	7.01	4.92	4.08	3.60	3.29	3.07	2.91	2.77	2.67	2.59	2.51	2.45	2.35	2.28	2.15	2.07	1.98	1.88	1.82	1.74	1.69	1.62	1.56	1.53	
80	3.96	3.11	2.72	2.48	2.33	2.21	2.12	2.05	1.99	1.95	1.91	1.88	1.82	1.77	1.70	1.65	1.60	1.54	1.51	1.45	1.42	1.38	1.35	1.32	80
	6.96	4.88	4.04	3.56	3.25	3.04	2.87	2.74	2.64	2.55	2.48	2.41	2.32	2.24	2.11	2.03	1.94	1.84	1.78	1.70	1.65	1.57	1.52	1.49	
100	3.94	3.09	2.70	2.46	2.30	2.19	2.10	2.03	1.97	1.92	1.88	1.85	1.79	1.75	1.68	1.63	1.57	1.51	1.48	1.42	1.39	1.34	1.30	1.28	100
	6.90	4.82	3.98	3.51	3.20	2.99	2.82	2.69	2.59	2.51	2.43	2.36	2.26	2.19	2.06	1.98	1.89	1.79	1.73	1.64	1.59	1.51	1.46	1.43	
125	3.92	3.07	2.68	2.44	2.29	2.17	2.08	2.01	1.95	1.90	1.86	1.83	1.77	1.72	1.65	1.60	1.55	1.49	1.45	1.39	1.36	1.31	1.27	1.25	125
	6.84	4.78	3.94	3.47	3.17	2.95	2.79	2.65	2.56	2.47	2.40	2.33	2.23	2.15	2.03	1.94	1.85	1.75	1.68	1.59	1.54	1.46	1.40	1.37	
150	3.91	3.06	2.67	2.43	2.27	2.16	2.07	2.00	1.94	1.89	1.85	1.82	1.76	1.71	1.64	1.59	1.54	1.47	1.44	1.37	1.34	1.29	1.25	1.22	150
	6.81	4.75	3.91	3.44	3.14	2.92	2.76	2.62	2.53	2.44	2.37	2.30	2.20	2.12	2.00	1.91	1.83	1.72	1.66	1.56	1.51	1.43	1.37	1.33	
200	3.89	3.04	2.65	2.41	2.26	2.14	2.05	1.98	1.92	1.87	1.83	1.80	1.74	1.69	1.62	1.57	1.52	1.45	1.42	1.35	1.32	1.26	1.22	1.19	200
	6.76	4.71	3.88	3.41	3.11	2.90	2.73	2.60	2.50	2.41	2.34	2.28	2.17	2.09	1.97	1.88	1.79	1.69	1.62	1.53	1.48	1.39	1.33	1.28	
400	3.86	3.02	2.62	2.39	2.23	2.12	2.03	1.96	1.90	1.85	1.81	1.78	1.72	1.67	1.60	1.54	1.49	1.42	1.38	1.32	1.28	1.22	1.16	1.13	400
	6.70	4.66	3.83	3.36	3.06	2.85	2.69	2.55	2.46	2.37	2.29	2.23	2.12	2.04	1.92	1.84	1.74	1.64	1.57	1.47	1.42	1.32	1.24	1.19	
1000	3.85	3.00	2.61	2.38	2.22	2.10	2.02	1.95	1.89	1.84	1.80	1.76	1.70	1.65	1.58	1.53	1.47	1.41	1.36	1.30	1.26	1.19	1.13	1.08	1000
	6.66	4.62	3.80	3.34	3.04	2.82	2.66	2.53	2.43	2.34	2.26	2.20	2.09	2.01	1.89	1.81	1.71	1.61	1.54	1.44	1.38	1.28	1.19	1.11	
∞	3.84	2.99	2.60	2.37	2.21	2.09	2.01	1.94	1.88	1.83	1.79	1.75	1.69	1.64	1.57	1.52	1.46	1.40	1.35	1.28	1.24	1.17	1.11	1.00	∞
	6.64	4.60	3.78	3.32	3.02	2.80	2.64	2.51	2.41	2.32	2.24	2.18	2.07	1.99	1.87	1.79	1.69	1.59	1.52	1.41	1.36	1.25	1.15	1.00	

Reprinted from Statistical Methods by G. W. Snedecor and W. G. Cochran, by permission of Iowa State University Press, Ames, Iowa

111

4M0995679